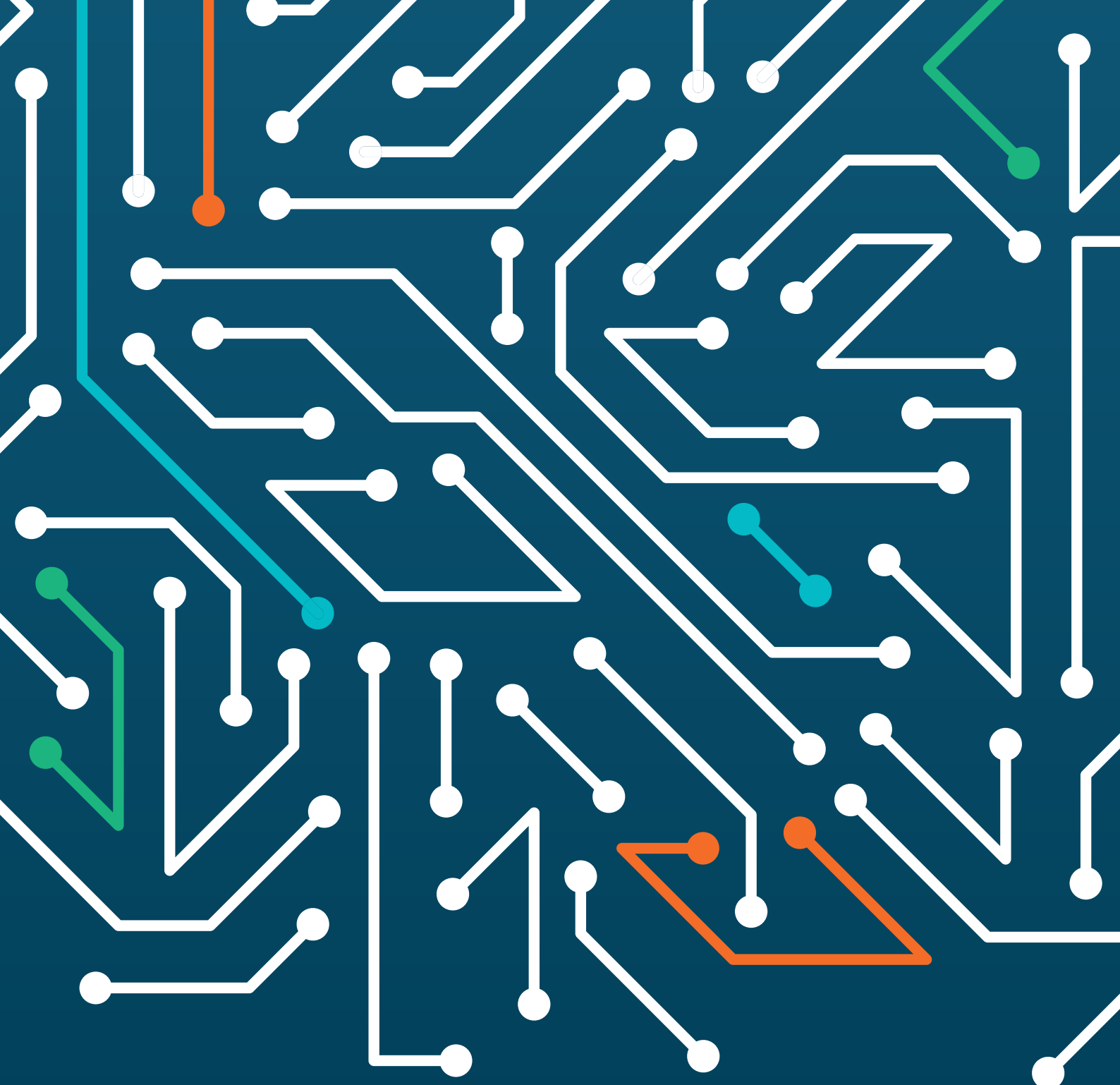


ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH

WHO GUIDANCE

World Health Organization

VISIT…

Ethics and governance of artificial intelligence for health: WHO guidance
ISBN 978-92-4-002920-0 (electronic version)
ISBN 978-92-4-002921-7 (print version)

Suggested citation. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

Cataloguing-in-Publication (CIP) data. CIP data are available at http://apps.who.int/iris.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see http://www.who.int/about/licensing.

Graphic design: The New Division

CONTENTS

FOREWORD

Ethics and Governance of Artificial Intelligence for Health
WHO Guidance
Foreword by Dr Soumya Swaminathan, Chief Scientist

"Our future is a race between the growing power of technology and the wisdom with which we use it."
Stephen Hawking

This quote by the famed physics Nobel Laureate reminds us of the great opportunities and challenges that new technologies hold in the health sector and beyond. In order to harness the power of science and innovation, WHO's Science Division was created in 2019 to support Member States in achieving the health-related Sustainable Development Goals (SDGs) and emergency preparedness and response. The Division provides global leadership in translating the latest in science, evidence, innovation, and digital solutions to improve health and health equity for all. This is in keeping with WHO's 13th Programme of Work (2019-2023) which stipulates that "...WHO's normative guidance will be informed by developments at the frontier of new scientific disciplines such as genomics, epigenetics, gene editing, artificial intelligence, and big data, all of which pose transformational opportunities but also risks to global health."

Artificial intelligence (AI) has enormous potential for strengthening the delivery of health care and medicine and helping all countries achieve universal health coverage. This includes improved diagnosis and clinical care, enhancing health research and drug development and assisting with the deployment of different public health interventions, such as disease surveillance, outbreak response, and health systems management.

AI could also benefit low- and middle-income countries, especially in countries that may have significant gaps in health care delivery and services for which AI could play a role. With the help of AI-based tools, governments could extend health care services to underserved populations, improve public health surveillance, and enable healthcare providers to better attend to patients and engage in complex care.

At the same time, for AI to have a beneficial impact on public health and medicine, ethical considerations and human rights must be placed at the centre of the design, development, and deployment of AI technologies for health. For AI to be used effectively for health, existing biases in healthcare services and systems based on race, ethnicity, age, and gender, that are encoded in data used to train algorithms, must be overcome. Governments will need to eliminate a pre-existing digital divide (or the uneven distribution of access to) the use of information and communication technologies. Such a digital divide not only limits use of AI in low- and middle-income countries but can also lead to the exclusion of populations in rich countries, whether based on gender, geography, culture, religion, language, or age.

Many of the world's largest technology companies are investing heavily in the collection of data (including health data), the development of algorithms, and AI deployment. The proliferation of AI could lead to the delivery of healthcare services in unregulated contexts and by unregulated providers, which

might create challenges for government oversight of health care. Therefore, appropriate regulatory oversight mechanisms must be developed to make the private sector accountable and responsive to those who can benefit from AI products and services, and can ensure that private sector decision-making and operations are transparent.

If employed wisely, AI has the potential to empower patients and communities to assume control of their own health care and better understand their evolving needs. But if we do not take appropriate measures, AI could also lead to situations where decisions that should be made by providers and patients are transferred to machines, which would undermine human autonomy, as humans may neither understand how an AI technology arrives at a decision, nor be able to negotiate with a technology to reach a shared decision. In the context of AI for health, autonomy means that humans should remain in full control of health-care systems and medical decisions.

This WHO guidance document is the result of a two-year development process led by two Departments in the Science Division - Digital Health and Innovation and Research For Health. WHO has worked with a leading group of twenty experts to identify core principles to promote the ethical use of AI for health - these are the first consensus principles in this field. The six core principles identified by the WHO Expert Group are the following: (1) Protect autonomy; (2) Promote human well-being, human safety, and the public interest; (3) Ensure transparency, explainability, and intelligibility; (4) Foster responsibility and accountability; (5) Ensure inclusiveness and equity; (6) Promote AI that is responsive and sustainable.

To implement these principles and human rights obligations into practice, all stakeholders, whether designers and programmers, providers, and patients, as well as Ministries of Health and Ministries of Information Technology, must work together to integrate ethical norms at every stage of a technology's design, development, and deployment.

Finally, I would like to thank all experts, stakeholders, and partners in the UN family and beyond who made essential contributions to the development of this document. I hope that this report will help to ensure that the development and use of AI for health will be guided by appropriate ethical norms and standards, so all populations can equally benefit from the great promise of these technologies in the future.

Dr Soumya Swaminathan

ACKNOWLEDGEMENTS

Development of this guidance document was led by Andreas Reis (Co-Lead, Health Ethics and Governance Unit, department of Research for Health) and Sameer Pujari (department of Digital Health and Innovation), under the overall guidance of John Reeder (Director, Research for Health), Bernardo Mariano (Director, Digital Health and Innovation) and Soumya Swaminathan (Chief Scientist).

Rohit Malpani (consultant, France) was the lead writer. The Co-Chairs of the Expert Group, Effy Vayena (ETH Zurich, Switzerland) and Partha Majumder (National Institute of Biomedical Genomics, India), provided overall guidance for the drafting of the report.

WHO is grateful to the following individuals who contributed to development of this guidance.

External expert group

Najeeb Al Shorbaji, eHealth Development Association, Jordan
Arisa Ema, Institute for Future Initiatives, The University of Tokyo, Japan
Amel Ghoulia, H3Africa, H3ABioNet, Tunisia
Jennifer Gibson, Joint Centre for Bioethics, Dalla Lana School of Public Health, University of Toronto, Canada
Kenneth W. Goodman, Institute for Bioethics and Health Policy, University of Miami Miller School of Medicines, USA
Jeroen van den Hoven, Delft University of Technology, Netherlands
Malavika Jayaram, Digital Asia Hub, Singapore
Daudi Jjingo, Makerere University, Uganda
Tze Yun Leong, National University of Singapore, Singapore
Alex John London, Carnegie Mellon University, USA
Partha Majumder, National Institute of Biomedical Genomics, India
Tshilidzi Marwala, University of Johannesburg, South Africa
Roli Mathur, Indian Council of Medical Research, India
Timo Minssen, Centre for Advanced Studies in Biomedical Innovation Law (CeBIL), Faculty of Law, University of Copenhagen, Denmark
Andrew Morris, Health Data Research UK, United Kingdom of Great Britain and Northern Ireland
Daniela Paolotti, ISI Foundation, Italy
Maria Paz Canales, Derechos Digitales, Chile
Jerome Singh, University of Kwa-Zulu Natal, South Africa
Effy Vayena, ETH Zurich, Switzerland
Robyn Whittaker, University of Auckland, New Zealand
Yi Zeng, Chinese Academy of Sciences, China

Observers

Tee Wee Ang, United Nations Educational, Scientific and Cultural Organization, France
Abdoulaye Banire Diallo, University of Quebec at Montreal, Canada
Julien Durand, International Federation of Pharmaceutical Manufacturers & Associations (IFPMA), Switzerland
David Gruson, Jouve, France
Lee Hibbard, Council of Europe, France
Lauren Milner, US Food and Drug Administration, USA
Rasha Abdul Rahim, Amnesty Tech, United Kingdom
Elettra Ronchi, Organization for Economic Co-operation and Development, France

External reviewers

Anurag Aggarwal, Council of Scientific and Industrial Research, India
Paolo Alcini, European Medicines Agency, Netherlands
Pamela Andanda, University of Witwatersrand, South Africa
Eva Blum-Dumontet, Privacy International, United Kingdom
Marcelo Corrales Compagnucci, CeBIL, Faculty of Law, University of Copenhagen, Denmark
Sara Leila Meg Davis, Graduate Institute, Switzerland
Juan M. Duran, Delft University of Technology, Netherlands
Osama El-Hassan, Dubai Health Authority, United Arab Emirates
Tomaso Falchetta, Privacy International, United Kingdom
Sara Gerke, Harvard Law School, USA
Tabitha Ha, STOP AIDS, United Kingdom
Henry Hoffman, ADA Health, Germany
Calvin Ho, University of Hong Kong, China, Hong Kong SAR
Prageeth Jayathissa, Vector Ltd, New Zealand
Otmar Kloiber, World Medical Association, France
Paulette Lacroix, International Medical Informatics Association, Canada
Hannah Yee-Fen Lim, Nanyang Technological University, Singapore
Allan Maleche, Kenya Legal and Ethical Issues Network on HIV and AIDS, Kenya
Peter Micek, Access Now, USA
Thomas Neumark, University of Oslo, Norway
Laura O'Brien, Access Now, USA
Alexandrine Pirlot de Corbion, Privacy International, United Kingdom
Léonard Van Rompaey, University of Copenhagen, Denmark
Tony Joakim Sandset, University of Oslo, Norway
Jay Shaw, Women's College Hospital Institute for Health System Solutions and Virtual Care, Canada
Sam Smith, medConfidential, United Kingdom
David Stewart, International Council of Nurses, Switzerland

External presenters at expert meetings

David Barbe, World Medical Association, USA
Elisabeth Bohn, Academy of Medical Sciences, United Kingdom
Katherine Chou, Google, USA
I. Glenn Cohen, Harvard Law School, USA
Naomi Lee, The Lancet, United Kingdom
Nada Malou, Médecins Sans Frontières, France
Vasantha Muthuswamy, Indian Council of Medical Research (retired), India
Sharon Kaur, A/P Gurmukh Singh, University of Malaya, Malaysia
Christian Stammel, Wearable Technologies, Germany
Alex Wang, Tencent, China
Kirstie Whitaker, Turing Institute, United Kingdom
Thomas Wiegand, Fraunhofer Heinrich Hertz Institute, Germany

WHO staff

Onyema Ajuebor, Technical Officer, Health Workforce, Geneva
Shada Alsalamah, Consultant, Digital Health and Innovation, Geneva
Ryan Dimentberg, Intern, Health Ethics and Governance Unit, Geneva
Clayton Hamilton, Technical Officer, WHO Regional Office for Europe, Copenhagen
Katherine Littler, Co-Lead, Health Ethics and Governance Unit, Geneva
Rohit Malpani, Consultant, Health Ethics and Governance Unit, Geneva
Ahmed Mohamed Amin Mandil, Coordinator, Research and Innovation, WHO Regional Office for the Eastern Mediterranean, Cairo
Bernardo Mariano, Chief Information Officer, Geneva
Issa T. Matta, Legal Affairs, Geneva
Vasee Moorthy, Coordinator, Research for Health, Geneva
Mohammed Hassan Nour, Technical Officer, Digital Health and Innovation, WHO Regional Office for the Eastern Mediterranean, Cairo
Lee-Anne Pascoe, Consultant, Health Ethics and Governance Unit, Geneva
Sameer Pujari, Technical Officer, Digital Health and Innovation, Geneva
John Reeder, Director, Research for Health, Geneva
Andreas Reis, Co-Lead, Health Ethics and Governance Unit, Geneva
John Reeder, Director, Research for Health, Geneva
Soumya Swaminathan, Chief Scientist, Geneva
Mariam Shokralla, Consultant, Digital Health and Innovation, Geneva
Diana Zandi, Technical Officer, Integrated Health Services, Geneva
Yu Zhao, Technical Officer, Digital Health and Innovation, Geneva

ABBREVIATIONS AND ACRONYMS

AI	artificial intelligence
CeBIL	Centre for Advanced Studies in Biomedical Innovation Law
EU	European Union
GDPR	General Data Protection Regulation
HIC	high-income countries
IP	intellectual property
LMIC	low- and middle-income countries
NHS	National Health Service (United Kingdom)
OECD	Organization for Economic Co-operation and Development
PPP	private–public partnership
SOFA	Sequential Organ Failure Assessment
UNESCO	United Nations Economic, Scientific and Cultural Organization
US	United States (of America)
USA	United States of America

EXECUTIVE SUMMARY

Artificial Intelligence (AI) refers to the ability of algorithms encoded in technology to learn from data so that they can perform automated tasks without every step in the process having to be programmed explicitly by a human. WHO recognizes that AI holds great promise for the practice of public health and medicine. WHO also recognizes that, to fully reap the benefits of AI, ethical challenges for health care systems, practitioners and beneficiaries of medical and public health services must be addressed. Many of the ethical concerns described in this report predate the advent of AI, although AI itself presents a number of novel concerns.

Whether AI can advance the interests of patients and communities depends on a collective effort to design and implement ethically defensible laws and policies and ethically designed AI technologies. There are also potential serious negative consequences if ethical principles and human rights obligations are not prioritized by those who fund, design, regulate or use AI technologies for health. AI's opportunities and challenges are thus inextricably linked.

AI can augment the ability of health-care providers to improve patient care, provide accurate diagnoses, optimize treatment plans, support pandemic preparedness and response, inform the decisions of health policy-makers or allocate resources within health systems. To unlock this potential, health-care workers and health systems must have detailed information on the contexts in which such systems can function safely and effectively, the conditions necessary to ensure reliable, appropriate use, and the mechanisms for continuous auditing and assessment of system performance. Health-care workers and health systems must have access to education and training in order to use and maintain these systems under the conditions for their safe, effective use.

AI can also empower patients and communities to assume control of their own health care and better understand their evolving needs. To achieve this, patients and communities require assurance that their rights and interests will not be subordinated to the powerful commercial interests of technology companies or the interests of governments in surveillance and social control. It also requires that the potential of AI to detect risks to patient or community health is incorporated into health systems in a way that advances human autonomy and dignity and does not displace humans from the centre of health decision-making.

AI can enable resource-poor countries, where patients often have restricted access to health-care workers or medical professionals, to bridge gaps in access to health services. AI systems must be carefully designed to reflect the diversity of socio-economic and health-care settings and be accompanied by training in digital skills, community engagement and awareness-raising. Systems based primarily on data of

individuals in high-income countries may not perform well for individuals in low- and middle-income settings. Country investments in AI and the supporting infrastructure should therefore help to build effective health-care systems by avoiding AI that encodes biases that are detrimental to equitable provision of and access to health-care services.

This guidance document, produced jointly by WHO's Health Ethics and Governance unit in the department of Research for Health and by the department of Digital Health and Innovation, is based on the collective views of a WHO Expert Group on Ethics and Governance of AI for Health, which comprised 20 experts in public health, medicine, law, human rights, technology and ethics. The group analysed many opportunities and challenges of AI and recommended policies, principles and practices for ethical use of AI for health and means to avoid its misuse to undermine human rights and legal obligations.

AI for health has been affected by the COVID-19 pandemic. Although the pandemic is not a focus of this report, it has illustrated the opportunities and challenges associated with AI for health. Numerous new applications have emerged for responding to the pandemic, while other applications have been found to be ineffective. Several applications have raised ethical concerns in relation to surveillance, infringement on the rights of privacy and autonomy, health and social inequity and the conditions necessary for trust and legitimate uses of data-intensive applications. During their deliberations on this report, members of the expert group prepared interim WHO guidance for the use of proximity tracking applications for COVID-19 contact-tracing.

Key ethical principles for the use of AI for health

This report endorses a set of key ethical principles. WHO hopes that these principles will be used as a basis for governments, technology developers, companies, civil society and inter-governmental organizations to adopt ethical approaches to appropriate use of AI for health. The six principles are summarized below and explained in depth in Section 5.

Protecting human autonomy: Use of AI can lead to situations in which decision-making power could be transferred to machines. The principle of autonomy requires that the use of AI or other computational systems does not undermine human autonomy. In the context of health care, this means that humans should remain in control of health-care systems and medical decisions. Respect for human autonomy also entails related duties to ensure that providers have the information necessary to make safe, effective use of AI systems and that people understand the role that such systems play in their care. It also requires protection of privacy and confidentiality and obtaining valid informed consent through appropriate legal frameworks for data protection.

Promoting human well-being and safety and the public interest. AI technologies should not harm people. The designers of AI technologies should satisfy regulatory requirements for safety, accuracy and efficacy for well-defined use cases or indications. Measures of quality control in practice and quality improvement in the use of AI over time should be available. Preventing harm requires that AI not result in mental or physical harm that could be avoided by use of an alternative practice or approach.

Ensuring transparency, explainability and intelligibility. AI technologies should be intelligible or understandable to developers, medical professionals, patients, users and regulators. Two broad approaches to intelligibility are to improve the transparency of AI technology and to make AI technology explainable. Transparency requires that sufficient information be published or documented before the design or deployment of an AI technology and that such information facilitate meaningful public consultation and debate on how the technology is designed and how it should or should not be used. AI technologies should be explainable according to the capacity of those to whom they are explained.

Fostering responsibility and accountability. Humans require clear, transparent specification of the tasks that systems can perform and the conditions under which they can achieve the desired performance. Although AI technologies perform specific tasks, it is the responsibility of stakeholders to ensure that they can perform those tasks and that AI is used under appropriate conditions and by appropriately trained people. Responsibility can be assured by application of "human warranty", which implies evaluation by patients and clinicians in the development and deployment of AI technologies. Human warranty requires application of regulatory principles upstream and downstream of the algorithm by establishing points of human supervision. If something goes wrong with an AI technology, there should be accountability. Appropriate mechanisms should be available for questioning and for redress for individuals and groups that are adversely affected by decisions based on algorithms.

Ensuring inclusiveness and equity. Inclusiveness requires that AI for health be designed to encourage the widest possible appropriate, equitable use and access, irrespective of age, sex, gender, income, race, ethnicity, sexual orientation, ability or other characteristics protected under human rights codes. AI technology, like any other technology, should be shared as widely as possible. AI technologies should be available for use not only in contexts and for needs in high-income settings but also in the contexts and for the capacity and diversity of LMIC. AI technologies should not encode biases to the disadvantage of identifiable groups, especially groups that are already marginalized. Bias is a threat to inclusiveness and equity, as it can result in a departure, often arbitrary, from equal treatment. AI technologies should minimize inevitable disparities in power that arise between providers and patients, between policy-makers and people and between companies and governments that create

and deploy AI technologies and those that use or rely on them. AI tools and systems should be monitored and evaluated to identify disproportionate effects on specific groups of people. No technology, AI or otherwise, should sustain or worsen existing forms of bias and discrimination.

Promoting AI that is responsive and sustainable. Responsiveness requires that designers, developers and users continuously, systematically and transparently assess AI applications during actual use. They should determine whether AI responds adequately and appropriately and according to communicated, legitimate expectations and requirements. Responsiveness also requires that AI technologies be consistent with wider promotion of the sustainability of health systems, environments and workplaces. AI systems should be designed to minimize their environmental consequences and increase energy efficiency. That is, use of AI should be consistent with global efforts to reduce the impact of human beings on the Earth's environment, ecosystems and climate. Sustainability also requires governments and companies to address anticipated disruptions in the workplace, including training for health-care workers to adapt to the use of AI systems, and potential job losses due to use of automated systems.

Overview of the report

This report is divided into nine sections and an annex. **Section 1** explains the rationale for WHO's engagement in this topic and the intended readership of the report's findings, analyses and recommendations. **Sections 2 and 3** define AI for health through its methods and applications. Section 2 provides a non-technical definition of AI, which includes several forms of machine learning as a subset of AI techniques. It also defines "big data," including sources of data that comprise biomedical or health big data. **Section 3** provides a non-comprehensive classification and examples of AI technologies for health, including applications used in LMIC, such as for medicine, health research, drug development, health systems management and planning, and public health surveillance.

Section 4 summarizes the laws, policies and principles that apply or could apply to the use of AI for health. These include human rights obligations as they apply to AI, the role of data protection laws and frameworks and other health data laws and policies. The section describes several frameworks that commend ethical principles for the use of AI for health, as well as the roles of bioethics, law, public policy and regulatory frameworks as sources of ethical norms.

Section 5 describes the six ethical principles that the Expert Group identified as guiding the development and use of AI for health. **Section 6** presents the ethical challenges identified and discussed by the Expert Group to which these guiding ethical principles can be applied: whether AI should be used; AI and the digital divide;

data collection and use; accountability and responsibility for decision-making with AI; autonomous decision-making; bias and discrimination associated with AI; risks of AI to safety and cybersecurity; impacts of AI on labour and employment in health care; challenges in the commercialization of AI for health care; and AI and climate change.

The final sections of the report identify legal, regulatory and non-legal measures for promoting ethical use of AI for health, including appropriate governance frameworks. Recommendations are provided.

Section 7 examines how various stakeholders can introduce ethical practices, programmes and measures to anticipate or meet ethical norms and legal obligations. They include: ethical, transparent design of AI technologies; mechanisms for the engagement and role of the public and demonstrating trustworthiness with providers and patients; impact assessment; and a research agenda for ethical use of AI for health care.

Section 8 is a discussion of how liability regimes may evolve with increasing use of AI for health care. It includes how liability could be assigned to a health-care provider, a technology provider and a health-care system or hospital that selects an AI technology and how the rules of liability might influence how a practitioner uses AI. The section also considers whether machine-learning algorithms are products, how to compensate individuals harmed by AI technologies, the role of regulatory agencies and specific aspects for LMIC.

Section 9 presents elements of a governance framework for AI for health. "Governance in health" refers to a range of functions for steering and rule-making by governments and other decision-makers, including international health agencies, to achieve national health policy objectives conducive to universal health coverage. The section analyses several governance frameworks either being developed or already matured. The frameworks discussed are: governance of data, control and benefit-sharing, governance of the private sector, governance of the public sector, regulatory considerations, the role of a policy observatory and model legislation and global governance of AI.

Finally, the report provides practical advice for implementing the WHO guidance for three sets of stakeholders: AI technology developers, ministries of health and health-care providers. The considerations are intended only as a starting-point for context-specific discussions and decisions by diverse stakeholders.

While the primary readership of this guidance document is ministries of health, it is also intended for other government agencies, ministries that will regulate AI, those who use AI technologies for health and entities that design and finance AI technologies for health.

Implementation of this guidance will require collective action. Companies and governments should introduce AI technologies only to improve the human condition and not for objectives such as unwarranted surveillance or to increase the sale of unrelated commercial goods and services. Providers should demand appropriate technologies and use them to maximize both the promise of AI and clinicians' expertise. Patients, community organizations and civil society should be able to hold governments and companies to account, to participate in the design of technologies and rules, to develop new standards and approaches and to demand and seek transparency to meet their own needs as well as those of their communities and health systems.

AI for health is a fast-moving, evolving field, and many applications, not yet envisaged, will emerge with ever-greater public and private investment. WHO may consider issuing specific guidance for additional tools and applications and may update this guidance periodically to keep pace with this rapidly changing field.

1. INTRODUCTION

Digital technologies and artificial intelligence (AI), particularly machine learning, are transforming medicine, medical research and public health. Technologies based on AI are now used in health services in countries of the Organization for Economic Co-operation and Development (OECD), and its utility is being assessed in low- and middle-income countries (LMIC). The United Nations Secretary-General has stated that safe deployment of new technologies, including AI, can help the world to achieve the United Nations Sustainable Development Goals (*1*), which would include the health-related objectives under Sustainable Development Goal 3. AI could also help to meet global commitments to achieve universal health coverage.

Use of AI for health nevertheless raises trans-national ethical, legal, commercial and social concerns. Many of these concerns are not unique to AI. The use of software and computing in health care has challenged developers, governments and providers for half a century, and AI poses additional, novel ethical challenges that extend beyond the purview of traditional regulators and participants in health-care systems. These ethical challenges must be adequately addressed if AI is to be widely used to improve human health, to preserve human autonomy and to ensure equitable access to such technologies.

Use of AI technologies for health holds great promise and has already contributed to important advances in fields such as drug discovery, genomics, radiology, pathology and prevention. AI could assist health-care providers in avoiding errors and allow clinicians to focus on providing care and solving complex cases. The potential benefits of these technologies and the economic and commercial potential of AI for health care presage ever greater use of AI worldwide.

Unchecked optimism in the potential benefits of AI could, however, veer towards habitual first recourse to technological solutions to complex problems. Such "techno-optimism" could make matters worse, for example, by exacerbating the unequal distribution of access to health-care technologies within and among wealthy and low-income countries (*2*). Furthermore, the digital divide could exacerbate inequitable access to health-care technologies by geography, gender, age or availability of devices, if countries do not take appropriate measures. Inappropriate use of AI could also perpetuate or exacerbate bias. Use of limited, low-quality, non-representative data in AI could perpetuate and deepen prejudices and disparities in health care. Biased inferences, misleading data analyses and poorly designed health applications and tools could be harmful. Predictive algorithms based on inadequate or inappropriate data can result in significant racial or ethnic bias. Use of high-quality, comprehensive datasets is essential.

AI could present a singular opportunity to augment and improve the capabilities of over-stretched health-care workers and providers. Yet, the introduction of AI for health care, as in many other sectors of the global economy, could have a significant negative impact on the health-care workforce. It could reduce the size of the workforce, limit, challenge or degrade the skills of health workers, and oblige them to retrain to adapt to the use of AI. Centuries of medical practice are based on relationships between provider and patient, and particular care must be taken when introducing AI technologies so that they do not disrupt such relationships.

The Universal Declaration of Human Rights, which includes pillars of patient rights such as dignity, privacy, confidentiality and informed consent, might be dramatically redefined or undermined as digital technologies take hold and expand. The performance of AI depends (among other factors) on the nature, type and volume of data and associated information and the conditions under which such data were gathered. The pursuit of data, whether by government or companies, could undermine privacy and autonomy at the service of government or private surveillance or commercial profit. If privacy and autonomy are not assured, the resulting limitation of the ability to exercise the full range of human rights, including civil and political rights (such as freedom of movement and expression) and social and economic rights (such as access to health care and education), might have a wider impact.

AI technologies, like many information technologies used in health care, are usually designed by companies or through public–private partnerships (PPPs), although many governments also develop and deploy these technologies. Some of the world's largest technology companies are developing new applications and services, which they either own or invest in. Many of these companies have already accumulated large quantities of data, including health data, and exercise significant power in society and the economy. While these companies may offer innovative approaches, there is concern that they might eventually exercise too much power in relation to governments, providers and patients.

AI technologies are also changing where people access health care. AI technologies for health are increasingly distributed outside regulated health-care settings, including at the workplace, on social media and in the education system. With the rapid proliferation and evolving uses of AI for health care, including in response to the COVID-19 pandemic, government agencies, academic institutions, foundations, nongovernmental organizations and national ethics committees are defining how governments and other entities should use and regulate such technologies effectively. Ethically optimized tools and applications could sustain widespread use of AI to improve human health and the quality of life, while mitigating or eliminating many risks and bad practices.

To date, there is no comprehensive international guidance on use of AI for health in accordance with ethical norms and human rights standards. Most countries do not have

laws or regulations to regulate use of AI technologies for health care, and their existing laws may not be adequate or specific enough for this purpose. WHO recognizes that ethics guidance based on the shared perspectives of the different entities that develop, use or oversee such technologies is critical to build trust in these technologies, to guard against negative or erosive effects and to avoid the proliferation of contradictory guidelines. Harmonized ethics guidance is therefore essential for the design and implementation of AI for global health.

The primary readership of this guidance document is ministries of health, as it is they that determine how to introduce, integrate and harness these technologies for the public good while restricting or prohibiting inappropriate use. The development, adoption and use of AI nevertheless requires an integrated, coordinated approach among government ministries beyond that for health. The stakeholders also include regulatory agencies, which must validate and define whether, when and how such technologies are to be used, ministries of education that teach current and future health-care workforces how such technologies function and are to be integrated into everyday practice, ministries of information technology that should facilitate the appropriate collection and use of health data and narrow the digital divide and countries' legal systems that should ensure that people harmed by AI technologies can seek redress.

This guidance document is also intended for the stakeholders throughout the health-care system who will have to adapt to and adopt these technologies, including medical researchers, scientists, health-care workers and, especially, patients. Access to such technologies can empower people who fall ill but can also leave them vulnerable, with fewer services and less protection. People have always been at the centre at all levels of decision-making in health care, whereas the inevitable growth of AI for health care could eventually challenge human primacy over medicine and health.

This guidance is also designed for those responsible for the design, deployment and refinement of AI technologies, including technologists and software developers. Finally, it is intended to guide the companies, universities, medical associations and international organizations that will, with governments and ministries of health, set policies and practices to define use of AI in the health sector. In identifying the many ethical concerns raised by AI and by providing the relevant ethical frameworks to address such concerns, this document is intended to support responsible use of AI worldwide.

WHO recognizes that AI is a fast-moving, evolving field and that many applications, not yet envisaged, will emerge as ever-greater public and private investment is dedicated to the use of AI for health. For example, in 2020, WHO issued interim guidance on the use of proximity tracking applications intended to facilitate contact-tracing during the COVID-19 pandemic. WHO may consider specific guidance for additional tools and applications and periodically update this guidance to keep pace with this rapidly changing field.

2. ARTIFICIAL INTELLIGENCE

"Artificial intelligence" generally refers to the performance by computer programs of tasks that are commonly associated with intelligent beings. The basis of AI is algorithms, which are translated into computer code that carries instructions for rapid analysis and transformation of data into conclusions, information or other outputs. Enormous quantities of data and the capacity to analyse such data rapidly fuel AI (*3*). A specific definition of AI in a recommendation of the Council on Artificial Intelligence of the OECD (*4*) states:

> An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

The various types of AI technology include machine-learning applications such as pattern recognition, natural language processing, signal processing and expert systems. Machine learning, which is a subset of AI techniques, is based on use of statistical and mathematical modelling techniques to define and analyse data. Such learned patterns are then applied to perform or guide certain tasks and make predictions.

Machine learning can be subcategorized according to how it learns from data into supervised learning, unsupervised learning and reinforced learning. In supervised learning, data used to train the model are labelled (the outcome variable is known), and the model infers a function from the data that can be used for predicting outputs from different inputs. Unsupervised learning does not involve labelling data but involves identification of hidden patterns in the data by a machine. Reinforcement learning involves machine learning by trial and error to achieve an objective for which the machine is "rewarded" or "penalized", depending on whether its inferences reach or hinder achievement of an objective (*5*). Deep learning, also known as "deep structured learning", is a family of machine learning based on use of multi-layered models to progressively extract features from data. Deep learning can be supervised, unsupervised or semi-supervised. Deep learning generally requires large amounts of data to be fed into the model.

Many machine-learning approaches are data-driven. They depend on large amounts of accurate data, referred to as "big data", to produce tangible results. "Big data" are complex data that are rapidly collected in such unprecedented quantities that terabytes (one trillion units [bytes] of digital information), petabytes (1000 terabytes)

or even zettabytes (one million petabytes) of storage space may be required as well as unconventional methods for their handling. The unique properties of big data are defined by four dimensions: volume, velocity, veracity and variety.

AI could improve the delivery of health care, such as prevention, diagnosis and treatment of disease (*6*), and is already changing how health services are delivered in several high-income countries (HIC). The possible applications of AI for health and medicine are expanding continually, although the use of AI may be limited outside HIC because of inadequate infrastructure. The applications can be defined according to the specific goals of use of AI and how AI is used to achieve those goals (methods). In health care, usable data have proliferated as a result of collection from numerous sources, including wearable technologies, genetic information generated by genome sequencing, electronic health-care records, radiological images and even from hospital rooms (*7*).

3. APPLICATIONS OF ARTIFICIAL INTELLIGENCE FOR HEALTH

This section identifies AI technologies developed and used in HIC, although examples of such technologies are emerging (and being pilot-tested or used) in LMIC. Digital health technologies are already used widely in LMIC, including for data collection, dissemination of health information by mobile phones and extended use of electronic medical records on open-software platforms and cloud computing (*8*). Schwabe and Wahl (*9*) have identified four uses of AI for health in LMIC: diagnosis, morbidity or mortality risk assessment, disease outbreaks and surveillance, and health policy and planning.

3.1 In health care

The use of AI in medicine raises notions of AI replacing clinicians and human decision-making. The prevailing sentiment is, however, that AI is increasingly improving diagnosis and clinical care, based on earlier definitions of the role of computers in medicine (*10*) and regulations in which AI is defined as a support tool (to improve judgement).

Diagnosis and prediction-based diagnosis

AI is being considered to support diagnosis in several ways, including in radiology and medical imaging. Such applications, while more widely used than other AI applications, are still relatively novel, and AI is not yet used routinely in clinical decision-making. Currently, AI is being evaluated for use in radiological diagnosis in oncology (thoracic imaging, abdominal and pelvic imaging, colonoscopy, mammography, brain imaging and dose optimization for radiological treatment), in non-radiological applications (dermatology, pathology), in diagnosis of diabetic retinopathy, in ophthalmology and for RNA and DNA sequencing to guide immunotherapy (*11*). In LMIC, AI may be used to improve detection of tuberculosis in a support system for interpreting staining images (*12*) or for scanning X-rays for signs of tuberculosis, COVID-19 or 27 other conditions (*13*).

Nevertheless, few such systems have been evaluated in prospective clinical trials. A recent comparison of deep-learning algorithms with health-care professionals in detection of diseases by medical imaging showed that AI is equivalent to human medical judgement in specific domains and applications in specific contexts but also that "few studies present externally validated results or compare the performance of deep learning models and health-care professionals using the same sample" (*14*). Other questions are whether the performance of AI can be generalized to implementation in practice and whether AI trained for use in one context can be used accurately and safely in a different geographical region or context.

As AI improves, it could allow medical providers to make faster, more accurate diagnoses. AI could be used for prompt detection of conditions such as stroke, pneumonia, breast cancer by imaging (*15*, *16*), coronary heart disease by echocardiography (*17*) and detection of cervical cancer (*18*). Unitaid, a United Nations agency for improving diagnosis and treatment of infectious diseases in LMIC, launched a partnership with the Clinton Health Access Initiative in 2018 to pilot-test use of an AI-based tool to screen for cervical cancer in India, Kenya, Malawi, Rwanda, South Africa and Zambia (*19*). Many low-income settings facing chronic shortages of health-care workers require assistance in diagnosis and assessment and to reduce their workload. It has been suggested that AI could fill gaps in the absence of health-care services or skilled workers *(9*).

AI might be used to predict illness or major health events before they occur. For example, an AI technology could be adapted to assess the relative risk of disease, which could be used for prevention of lifestyle diseases such as cardiovascular disease (*20*, *21*) and diabetes (*22*). Another use of AI for prediction could be to identify individuals with tuberculosis in LMIC who are not reached by the health system and therefore do not know their status (*23*). Predictive analytics could avert other causes of unnecessary morbidity and mortality in LMIC, such as birth asphyxia. An expert system used in LMIC is 77% sensitive and 95% specific for predicting the need for resuscitation (*8*). Several ethical challenges to prediction-based health care are discussed in section 6.5.

Clinical care

Clinicians might use AI to integrate patient records during consultations, identify patients at risk and vulnerable groups, as an aid in difficult treatment decisions and to catch clinical errors. In LMIC, for example, AI could be used in the management of antiretroviral therapy by predicting resistance to HIV drugs and disease progression, to help physicians optimize therapy (23). Yet, clinical experience and knowledge about patients is essential, and AI will not be a substitute for clinical due diligence for the foreseeable future. If it did, clinicians might engage in "automation bias" and not consider whether an AI technology meets their needs or those of the patient. (See section 6.4.)

The wider use of AI in medicine also has technological challenges. Although many prototypes developed in both the public and the private sectors have performed well in field tests, they often cannot be translated, commercialized or deployed. An additional obstacle is constant changes in computing and information technology management, whereby systems become obsolete ("software erosion") and companies disappear. In resource-poor countries, the lack of digital infrastructure and the digital divide (See section 6.2.) will limit use of such technologies.

Health-care workers will have to adapt their clinical practice significantly as use of AI increases. AI could automate tasks, giving doctors time to listen to patients, address their fears and concerns and ask about unrelated social factors, although they may still worry about their responsibility and accountability. Doctors will have to update their competence to communicate risks, make predictions and discuss trade-offs with patients and also express their ethical and legal concern about understanding AI technology. Even if technology makes the predicted gains, those gains will materialize only if the individuals who manage health systems use them to extend the capacity of the health system in other areas, such as better availability of medicines or other prescribed interventions or forms of clinical care.

Emerging trends in the use of AI in clinical care

Several important changes imposed by the use of AI in clinical care extend beyond the provider–patient relationship. Four trends described here are: the evolving role of the patient in clinical care; the shift from hospital to home-based care; use of AI to provide "clinical" care outside the formal health system; and use of AI for resource allocation and prioritization. Each of these trends has ethical implications, as discussed below.

The evolving role of the patient in clinical care

AI could eventually change how patients self-manage their own medical conditions, especially chronic diseases such as cardiovascular diseases, diabetes and mental problems (*24*). Patients already take significant responsibility for their own care, including taking medicines, improving their nutrition and diet, engaging in physical activity, caring for wounds or delivering injections. AI could assist in self-care, including through conversation agents (e.g. "chat bots"), health monitoring and risk prediction tools and technologies designed specifically for individuals with disabilities (24). While a shift to patient-based care may be considered empowering and beneficial for some patients, others might find the additional responsibility stressful, and it might limit an individual's access to formal health-care services.

The growing use of digital self-management applications and technologies also raises wider questions about whether such technologies should be regulated as clinical applications, thus requiring greater regulatory scrutiny, or as "wellness applications", requiring less regulatory scrutiny. Many digital self-management technologies arguably fall into a "grey zone" between these two categories and may present a risk if they are used by patients for their own disease management or clinical care but remain largely unregulated or could be used without prior medical advice. Such concerns are exacerbated by the distribution of such applications by entities that are not a part of the formal health-care system. This related but separate trend is discussed below.

The shift from hospital to home-based care

Telemedicine is part of a larger shift from hospital- to home-based care, with use of AI technologies to facilitate the shift. They include remote monitoring systems, such as video-observed therapy for tuberculosis and virtual assistants to support patient care. Even before the COVID-19 pandemic, over 50 health-care systems in the USA were making use of telemedicine services (*25*). COVID-19, having discouraged people in many settings from visiting health-care facilities, accelerated and expanded the use of telemedicine in 2020, and the trend is expected to continue. In China, the number of telemedicine providers has increased by nearly four times during the pandemic (*26*).

The shift to home-based care has also partly been facilitated by increased use of search engines (which rely on algorithms) for medical information as well as by the growth in the number of text or speech chatbots for health care (*27*), the performance of which has improved with improvements in natural language processing, a form of AI that enables machines to understand human language. The use of chatbots has also accelerated during the COVID-19 pandemic (*28*).

Furthermore, AI technologies may play a more active role in the management of patients' health outside clinical settings, such as in "just-in-time adaptive interventions". These rely on sensors to provide patients with specific interventions according to data collected previously and currently; they also notify a health-care provider of any emerging concern (*29*). The growth and use of sensors and wearables may improve the effectiveness of "just-in-time adaptive interventions" but also raise concern, in view of the amount of data such technologies are collecting, how they are used and the burden such technologies may shift to patients.

Use of AI to extend "clinical" care beyond the formal health-care system

AI applications in health are no longer exclusively used in health-care systems (or home care), as AI technologies for health can be readily acquired and used by non-health system entities. This has meant that people can now obtain health-care services outside the health-care system. For example, AI applications for mental health are often provided through the education system, workplaces and social media and may even be linked to financial services (*30*). While there may be support for such extended uses of health applications to compensate for both increased demand and a limited number of providers (31), they generate new questions and concerns. (See section 9.3.)

These three trends may require near-continuous monitoring (and self-monitoring) of people, even when they are not sick (or are "patients"). AI-guided technologies require the use of mobile health applications and wearables, and their use has increased with the trend to self-management (*31*). Wearable technologies include those placed in the body (artificial limbs, smart implants), on the body (insulin pump patches, electroencephalogram devices) or near the body (activity trackers, smart watches and

smart glasses). By 2025, 1.5 billion wearable units may be purchased annually. [1] Wearables will create more opportunities to monitor a person's health and to capture more data to predict health risks, often with greater efficiency and in a timelier manner.

Although such monitoring of "healthy" individuals could generate data to predict or detect health risks or improve a person's treatment when necessary, it raises concern, as it permits near-constant surveillance and collection of excessive data that otherwise should remain unknown or uncollected. Such data collection also contributes to the ever-growing practice of "biosurveillance", a form of surveillance for health data and other biometrics, such as facial features, fingerprints, temperature and pulse (*32*). The growth of biosurveillance poses significant ethical and legal concerns, including the use of such data for medical and non-medical purposes for which explicit consent might not have been obtained or the repurposing of such data for non-health purposes by a government or company, such as within criminal justice or immigration systems. (See section 6.3.) Thus, such data should be liable to the same levels of data protection and security as for data collected on an individual in a formal clinical care setting.

Use of AI for resource allocation and prioritization

AI is being considered for use to assist in decision-making about prioritization or allocation of scarce resources. Prognostic scoring systems have long been available in critical care units. One of the best-known, Sequential Organ Failure Assessment (SOFA) (*33*), for analysis of the severity of illness and for predicting mortality, has been in use for decades, and SOFA scores have been widely used in some jurisdictions to guide allocation of resources for COVID-19 (*34*). It is not an AI system; however, an AI version, "DeepSOFA" (*35*), has been developed.

The growing attraction of this use of AI has been due partly to the COVID-19 pandemic, as many institutions lack bed capacity and others have inadequate ventilators. Thus, hospitals and clinics in the worst-affected countries have been overwhelmed.
It has been suggested that machine-learning algorithms could be trained and used to assist in decisions to ration supplies, identify which individuals should receive critical care or when to discontinue certain interventions, especially ventilator support (*36*).
AI tools could also be used to guide allocation of other scarce health resources during the COVID-19 pandemic, such as newly approved vaccines for which there is an insufficient initial supply (*37*).

Several ethical challenges associated with the use of AI for resource allocation and prioritization are described in section 6.5.

[1] Presentation by Christian Stammel. Wearable Technologies, Germany, to the WHO Meeting of the Expert Group on Ethics and Governance of AI for Health, 6 March 2020.

3.2 In health research and drug development

Application of AI for health research

An important area of health research with AI is based on use of data generated for electronic health records. Such data may be difficult to use if the underlying information technology system and database do not discourage the proliferation of heterogeneous or low-quality data. AI can nevertheless be applied to electronic health records for biomedical research, quality improvement and optimization of clinical care. From electronic health records, AI that is accurately designed and trained with appropriate data can help to identify clinical best practices before the customary pathway of scientific publication, guideline development and clinical support tools. AI can also assist in analysing clinical practice patterns derived from electronic health records to develop new clinical practice models.

A second (of many) application of AI for health research is in the field of genomics. Genomics is the study of the entire genetic material of an organism, which in humans consists of an estimated three billion DNA base pairs. Genomic medicine is an emerging discipline based on individuals' genomic information to guide clinical care and personalized approaches to diagnosis and treatment (*38*). As the analysis of such large datasets is complex, AI is expected to play an important role in genomics. In health research, for example, AI could improve human understanding of disease or identify new disease biomarkers (*38*), although the quality of the data and whether they are representative and unbiased (See section 6.6.) could undermine the results.

Uses of AI in drug development

AI is expected in time to be used to both simplify and accelerate drug development. AI could change drug discovery from a labour-intensive to a capital- and data-intensive process with the use of robotics and models of genetic targets, drugs, organs, diseases and their progression, pharmacokinetics, safety and efficacy. AI could be used in drug discovery and throughout drug development to shorten the process and make it less expensive and more effective (*39*). AI was used to identify potential treatments for Ebola virus disease, although, as in all drug development, identification of a lead compound may not result in a safe, effective therapy (*40*).

In December 2020, DeepMind announced that its AlphaFold system had solved what is known as the "protein folding problem", in that the system can reliably predict the three-dimensional shape of a protein (*41*). Although this achievement is only one part of a long process in understanding diseases and developing new medicines and vaccines, it should help to speed the development of new medicines and improve the repurposing of existing medicines for use against new viruses and new diseases (*41*). While this advance could significantly accelerate drug discovery, there is ethical concern about ownership and control of an AI technology that could be critical to drug development, as it might eventually be available to government, not-for-profit, academic and LMIC researchers only under commercial terms and conditions that limit its diffusion and use.

At present, drug development is led either by humans or by AI with human oversight. In the next two decades, as work with machines is optimized, the role of AI could evolve. Computing is starting to facilitate drug discovery and development by finding novel leads and evaluating whether they meet the criteria for new drugs, structuring unorganized data from medical imaging, searching large volumes of data, including health-care records, genetics data, laboratory tests, the Internet of Things, published literature and other types of health big data to identify structures and features, while recreating the body and its organs on chips (tissue chips) for AI analysis (*39*, *42*). By 2040, testing of medicines might be virtual – without animals or humans – based on computer models of the human body, tumours, safety, efficacy, epigenetics and other parameters. Prescription drugs could be designed for each person. Such efforts could contribute to precision medicine or health care that is individually tailored to a person's genes, lifestyle and environment.

3.3 In health systems management and planning

Health systems, even in a single-payer, government-run system, may be overly complex and involve numerous actors who contribute to, pay for or benefit from the provision of health-care services. The management and administration of care may be laborious. AI can be used to assist personnel in complex logistical tasks, such as optimization of the medical supply chain, to assume mundane, repetitive tasks or to support complex decision-making. Some possible functions of AI for health systems management include: identifying and eliminating fraud or waste, scheduling patients, predicting which patients are unlikely to attend a scheduled appointment and assisting in identification of staffing requirements (*43*).

AI could also be useful in complex decision-making and planning, including in LMIC. For example, researchers in South Africa applied machine-learning models to administrative data to predict the length of stay of health workers in underserved communities (*9*). In a study in Brazil, researchers used several government data sets and AI to optimize the allocation of health-system resources by geographical location according to current health challenges (*9*). Allocation of scarce health resources through use of AI has raised concern, however, that resources may not be fairly allocated due, for example, to bias in the data. (See section 6.5.)

3.4 In public health and public health surveillance

Several AI tools for population and public health can be used in public health programmes. For example, new developments in AI could, after rigorous evaluation, improve identification of disease outbreaks and support surveillance. Several concerns about the use of technology for public health surveillance, promotion and outbreak response must, however, be considered before use of AI for such purposes, including the tension between the public health benefits of surveillance and ethical and legal concern about individual (or community) privacy and autonomy (*44*).

Health promotion

AI can be used for health promotion or to identify target populations or locations with "high-risk" behaviour and populations that would benefit from health communication and messaging (micro-targeting). AI programmes can use different forms of data to identify such populations, with varying accuracy, to improve message targeting. Micro-targeting can also, however, raise concern, such as that with respect to commercial and political advertising, including the opaqueness of processes that facilitate micro-targeting. Furthermore, users who receive such messages may have no explanation or indication of why they have been targeted (*45*). Micro-targeting also undermines a population's equal access to information, can affect public debate and can facilitate exclusion or discrimination if it is used improperly by the public or private sector.

Disease prevention

AI has also been used to address the underlying causes of poor health outcomes, such as risks related to environmental or occupational health. AI tools can be used to identify bacterial contamination in water treatment plants, simplify detection and lower the costs. Sensors can also be used to improve environmental health, such as by analysing air pollution patterns or using machine learning to make inferences between the physical environment and healthy behaviour (*29*). One concern with such use of AI is whether it is provided equitably or if such technologies are used only on behalf of wealthier populations and regions that have the relevant infrastructure for its use (*46*).

Surveillance (including prediction-based surveillance) and emergency preparedness

AI has been used in public health surveillance for collecting evidence and using it to create mathematical models to make decisions. Technology is changing the types of data collected for public health surveillance by the addition of digital "traces", which are data that are not generated specifically for public health purposes (such as from blogs, videos, official reports and Internet searches). Videos (e.g. YouTube) are another "rich" source of information for health insights (*47*).

Characterization of digital traces as "health data" raises questions about the types of privacy protection or other safeguards that should be attached to such datasets if they are not publicly available. For example, the use of digital traces as health data could violate the data protection principle of "purpose limitation", that individuals who generate such data should know what their data will be used for at the point of collection (*48*).

Such use also raises questions of accuracy. Models are useful only when appropriate data are used. Machine-learning algorithms could be more valuable when augmented by digital traces of human activity, yet such digital traces could also negatively impact an algorithm's performance. Google Flu Trends, for example, was based on search engine queries about complications, remedies, symptoms and antiviral medications for

influenza, which are used to estimate and predict influenza activity. While Google Flu Trends first provided relatively accurate predictions before those of the US Centers for Disease Control and Prevention, it overestimated the prevalence of flu between 2011 and 2013 because the system was not re-trained as human search behaviour evolved (*49*).

Although many public health institutions are not yet making full use of these sources of data, surveillance itself is changing, especially real-time surveillance. For example, researchers could detect a surge in cases of severe pulmonary disease associated with the use of electronic cigarettes by mining disparate online sources of information and using Health Map, an online data-mining tool (*50*). Similarly, Microsoft researchers have found early evidence of adverse drug reactions from web logs with an AI system. In 2013, the company's researchers detected side-effects of several prescription drugs before they were found by the US Food and Drug Administration's warning system (*51*). In 2020, the US Food and Drug Administration sponsored a "challenge", soliciting public submissions to develop computation algorithms for automatic detection of adverse events from publicly available data *(52*). Despite its potential benefits, real-time data collection, like the collection and use of digital traces, could violate data protection rules if surveillance was not the purpose of its initial collection, which is especially likely when data collection is automated.

Before the COVID-19 pandemic, WHO had started to develop EPI-BRAIN, a global platform that will allow experts in data and public health to analyse large datasets for emergency preparedness and response. (See also section 7.1.) AI has been used to assist in both detection and prediction during the COVID-19 pandemic, although some consider that the techniques and programming developed will "pay dividends" only during a subsequent pandemic (*49*). HealthMap first issued a short bulletin about a new type of pneumonia in Wuhan, China, at the end of December 2019 (*49*). Since then, AI has been used to "now-cast" (assess the current state of) the COVID-19 pandemic (*49*), while, in some countries, real-time data on the movement and location of people has been used to build AI models to forecast regional transmission dynamics and guide border checks and surveillance (*53*). In order to determine how such applications should be used, an assessment should be conducted of whether they are accurate, effective and useful.

Outbreak response

The possible uses of AI for different aspects of outbreak response have also expanded during the COVID-19 pandemic. They include studying SARS-CoV2 transmission, facilitating detection, developing possible vaccines and treatments and understanding the socio-economic impacts of the pandemic (*54*). Such use of AI was already tested during the pandemic of Ebola virus disease in West Africa in 2014, although the assumptions underlying use of AI technologies to predict the spread of the Ebola virus were based on erroneous views of how the virus was spreading (*55*, *56*). While many

possible uses of AI have been identified and used during the COVID-19 pandemic, their actual impact is likely to have been modest; in some cases, early AI screening tools for SARS-CoV2 "were utter junk" with which companies "were trying to capitalise on the panic and anxiety" (*57*).

New applications (*58*) are intended to support the off-line response, although not all may involve use of AI. These have included proximity tracking applications intended to notify users (and possibly health authorities) that they have been in the proximity (for some duration) of an individual who subsequently tested positive for SARS-CoV2. Concern has been raised about privacy and the utility and accuracy of proximity-tracking applications, and WHO issued interim guidance on the ethical use of proximity-tracking applications in 2020 (*59*).

WHO and many ministries of health have also deployed symptom checkers, which are intended to guide users through a series of questions to assist in determining whether they should seek additional medical advice or testing for SARS-CoV2. The first symptom checkers were "hard coded", based on accumulated clinical judgement, as there were no previous data, and on a simple decision tree from older AI techniques, which involved direct encoding of expert knowledge. AI systems based on machine learning require accurate training, while data are initially scarce for a new disease such as COVID-19 (*60*). New symptom checkers are based on machine learning to provide advice to patients (*61*), although their effectiveness is not yet known; all symptom checkers require that users provide accurate information.

AI has also been introduced to map the movements of individuals in order to approximate the effectiveness of government-mandated orders to remain in confinement, and, in some countries, AI technology has been used to identify individuals who should self-quarantine and be tested. These technologies raise legal and ethical concerns about privacy and risk of discrimination and also about possibly unnecessary restriction of movement or access to services, which heavily impact the exercise of a range of human rights (*53*). As for all AI technologies, their actual effectiveness depends on whether the datasets are representative of the populations in which the technologies are used, and they remain questionable without systematic testing and evaluation. The uses described above are therefore not yet established.

3.5 The future of artificial intelligence for health

While AI may not replace clinical decision-making, it could improve decisions made by clinicians. In settings with limited resources, AI could be used to conduct screening and evaluation if insufficient medical expertise is available, a common challenge in many resource-poor settings. Yet, whether AI can advance beyond narrow tasks depends on numerous factors beyond the state of AI science and on the trust of providers, patients and health-care professionals in AI-based technologies. In the

following sections of this report, ethical concerns and risks associated with the expanding use of AI for health are discussed, including by whom and how such technologies are deployed and developed. Technological, legal, security and ethical challenges and concerns are discussed not to dissuade potential use of AI for health but to ensure that AI fulfils its great potential and promise.

4. LAWS, POLICIES AND PRINCIPLES THAT APPLY TO ARTIFICIAL INTELLIGENCE FOR HEALTH

Laws, policies and principles for regulating and managing the use of AI and specifically use of AI for health are fragmented and limited. Numerous principles and guidelines have been developed for application of "ethical" AI in the private and public sectors and in research institutions (*62*); however, there is no consensus on its definition, best practices or ethical requirements, and different legal regimes and governance models are associated with each set of principles. Other norms, rules and frameworks also apply to use of AI, including human rights obligations, bioethics laws and policies, data protection laws and regulatory standards. These are summarized below and discussed elsewhere in the report. Section 5 provides a set of guiding principles agreed by the WHO Expert Group by consensus, on which this analysis and these findings are based.

4.1 Artificial intelligence and human rights

Efforts to enumerate human rights and to fortify their observance through explicit legal mechanisms are reflected in international and regional human rights conventions, including the Universal Declaration on Human Rights, the International Covenant on Economic, Social and Cultural Rights (including General Comment No. 14, which defines the right to health), the International Covenant on Civil and Political Rights and regional human rights conventions, such as the African Charter on Human and People's Rights, the American Convention on Human Rights and the European Convention on Human Rights. Not all governments have acceded to key human rights instruments; some have signed but not ratified such charters or have expressed reservations to certain provisions. In general, however, human rights listed in international instruments establish a baseline for the protection and promotion of human dignity worldwide and are enforced through national legislation such as constitutions or human rights legislation.

Machine-learning systems could advance human rights but could also undermine core human rights standards. The Office of the High Commissioner for Human Rights has issued several opinions on the relation of AI to the realization of human rights. In guidance issued in March 2020, the Office noted that AI and big data can improve the human right to health when "new technologies are designed in an accountable manner" and could ensure that certain vulnerable populations have efficient, individualized care, such as assistive devices, built-in environmental applications and robotics (*63*). The Office also noted, however, that such technologies could dehumanize care, undermine the autonomy and independence of older persons and pose significant risks to patient privacy – all of which are contrary to the right to health (*63*). In February 2021, in a speech to the Human Rights Council, the United Nations

Secretary-General noted a number of concerns for human rights associated with the growing collection and use of data on the COVID-19 pandemic and called on governments to "place human rights at the centre of regulatory frameworks and legislation on the development and use of digital technologies" (*64*). Human rights organizations have interpreted and, when necessary, adapted existing human rights laws and standards to AI assessment and are reviewing them in the face of the challenges and opportunities associated with AI. The Toronto Declaration (*65*) addresses the impact of AI on human rights and situates AI within the universally binding, actionable framework of human rights laws and standards; it provides mechanisms for public and private sector accountability and the protection of people from discrimination and promotes equity, diversity and inclusion, while safeguarding equality and effective redress and remedy.

In 2018, the Council of Europe's Committee of Ministers issued draft recommendations to Member States on the impact of algorithmic systems on human rights (*66*). The Council of Europe is further examining the feasibility and potential elements of a legal framework for the development, design and application of digital technologies according to its standards on human rights, democracy and the rule of law.
Legal frameworks for human rights, bioethics and privacy adopted by countries are applicable to several aspects of AI for health. They include Article 8 of the European Convention on Human Rights: the right to respect for private and family life, home and correspondence (*67*); the Oviedo Convention on Human Rights and Biomedicine, which covers ethical principles of individual human rights and responsibilities (68); the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (*69*) and guidelines on the protection of individuals with regard to the processing of personal data in a world of big data, prepared by the Consultative Committee of Convention 108+ (*69*).

Yet, even with robust human rights standards, organizations and institutions recognize that better definition is required of how human rights standards and safeguards relate and apply to AI and that new laws and jurisprudence are required to address the interaction of AI and human rights. New legal guidance has been prepared by the Council of Europe. In 2019–2020, the Council established the Ad-hoc Committee on Artificial Intelligence to conduct broad multi-stakeholder consultations in order to determine the feasibility and potential elements of a legal framework for the design and application of AI according to the Council of Europe's standards on human rights, democracy and the rule of law. Further, in 2019, the Council of Europe released Guidelines on artificial intelligence and data protection (*70*), also based on the protection of human dignity and safeguarding human rights and fundamental freedom. In addition, the ethical charter of the European Commission for Efficiency of Justice includes five principles relevant to use of AI for health (*71*).

4.2 Data protection laws and policies

Data protection laws are "rights-based approaches" that provide standards for regulating data processing that both protect the rights of individuals and establish obligations for data controllers and processors. Data protection laws also increasingly recognize that people have the right not to be subject to decisions guided solely by automated processes. Over 100 countries have enacted data protection laws. One well-known set of data protection laws is the General Data Protection Regulation (GDPR) of the European Union (EU); in the USA, the Health Insurance Portability and Accountability Act, enacted in 1996, applies to privacy and to the security of health data.

Some standards and guidelines are designed specifically to manage the use of personal data for AI. For example, the Ibero-American Data Protection Network, which consists of 22 data protection authorities in Portugal and Spain and in Mexico and other countries in Central and South America and the Caribbean, has issued General Recommendations for the Processing of Personal Data in Artificial Intelligence (72) and specific guidelines for compliance with the principles and rights that govern the protection of personal data in AI projects (*73*).

4.3 Existing laws and policies related to health data

Several types of laws and policies govern the collection, processing, analysis, transfer and use of health data. The Council of Europe's Committee of Ministers issued a recommendation to Member States on the protection of health-related data in 2019 (*74*), and the African Union's convention on cybersecurity and personal data protection (2014) (*75*) requires that personal data involving genetic information and health research be processed only with the authorization of the national data protection authority through the Personal Data Protection Guidelines for Africa (*76*). Generally, the African continent's digital transformation strategy (*77*) encourages African Union Member States to "have adequate regulation; particularly around data governance and digital platforms, to ensure that trust is preserved in the digitalization". In February 2021, the African Academy of Sciences and the African Union Development Agency released recommendations for data and biospecimen governance in Africa to promote a participant-centred approach to research involving human participants, while enabling ethical research practices on the continent and providing guidelines for governance (*78*).

Laws that govern the transfer of data among countries include those defined in trade agreements, intellectual property (IP) rules for the ownership of data and the role of competition law and policy related to the accumulation and control of data (including health data). These are discussed in detail later in this report.

4.4 General principles for the development and use of artificial intelligence

An estimated 100 proposals for AI principles have been published in the past decade, and studies have been conducted to identify which principles are most cited (*79*). In one study of mapping and analysis of current principles and guidelines for ethical use of AI, convergence was found on transparency, justice, fairness, non-maleficence and responsibility, while other principles such as privacy, solidarity, human dignity and sustainability were under-represented (*62*).

Several intergovernmental organizations and countries have proposed such principles (Box 1).

Box 1. Examples of AI ethics principles proposed by intergovernmental organizations and countries

- The Recommendations of the OECD Council on Artificial Intelligence (*80*), the first intergovernmental standard on AI, were adopted in May 2019 by OECD's 36 member countries and have since been applied by a number of partner economies. The OECD AI principles (*81*) provided the basis for the AI principles endorsed by G20 governments in June 2019 (*82*). While OECD recommendations are not legally binding, they carry a political commitment and have proved highly influential in setting international standards in other policy areas (e.g. privacy and data protection) and helping governments to design national legislation. The OECD launched an online platform for public policy on AI, the AI Policy Observatory (*83*) (See section 9.6.) and is cooperating on this and other initiatives on the ethical implications of AI with the Council of Europe, the United Nations Economic, Scientific and Cultural Organization (UNESCO) and WHO.
- In 2019, the Council of Europe Commissioner for Human Rights issued recommendations to ensure that human rights are strengthened rather than undermined by AI: Unboxing artificial intelligence: 10 steps to protect human rights recommendations (*84*).
- The European Commission appointed 52 representatives from academia, civil society and industry to its High-level Expert Group on Artificial Intelligence and issued Ethics Guidelines for Trustworthy AI (*85*).
- Japan has issued several guidelines on the use of AI, including on research and development and utilization (*86*).
- China has issued National Governance Principles for the New Generation Artificial Intelligence, which serves as the national principles for AI governance in China (*87*). Academia and industry have jointly issued the Beijing Artificial Intelligence Principles (*88*).[2]
- In Singapore, a series of initiatives on AI governance and ethics was designed to build an ecosystem of trust to support adoption of AI. They include Asia's first Model AI governance framework, released in January 2019; an international industry-led Advisory Council on the Ethical Use of AI and Data formed in June 2018; a research programme on the governance of AI and data use established in partnership with the Singapore Management University in September 2018 (*89*); and a certification programme for ethics and governance of AI for companies and developers (*90*).
- The African Union's High-level Panel on Emerging Technologies is preparing broad guidance on the use of AI to promote economic development and its use in various sectors, including health care (*91*).

[2] Presentation by Professor Yi Zeng, Chinese Academy of Sciences, 4 October 2019, to the WHO working group on ethics and governance of AI for health.

4.5 Principles for use of artificial intelligence for health

No specific ethical principles for use of AI for health have yet been proposed for adoption worldwide. Before WHO's work on guidance on the ethics and governance of AI for health, the WHO Global Conference on Primary Health Care issued the Astana Declaration (*92*), which includes principles for the use of digital technology. The Declaration calls for promotion of rational, safe use and protection of personal data and use of technology to improve access to health care, enrich health service delivery, improve the quality of service and patient safety and increase the efficiency and coordination of care.

UNESCO has guidance and principles for the use of AI in general and for the use of big data in health. UNESCO's work on the ethical implications of AI is supported by two standing expert committees, the World Commission on the Ethics of Scientific Knowledge and Technology and the International Bioethics Committee. Other work includes the report of the International Bioethics Committee on big data and health in 2017, which identified important elements of a governance framework (*93*); the World Commission on the Ethics of Scientific Knowledge and Technology report on robotics ethics in 2017 (*94*); a preliminary study on the ethics of AI by UNESCO in 2019, which raised ethical concern about education, science and gender (*95*); a recommendation on the ethics of AI to be considered by UNESCO's General Conference in 2021; and a report by the World Commission on the Ethics of Scientific Knowledge and Technology on the Internet of Things.

In 2019, the United Kingdom's National Health Service (NHS) released a code of conduct, with 10 principles for the development and use of safe, ethical, effective, data-based health and care technologies (*96*). In October 2019, The Lancet and The Financial Times launched a joint commission, The Governing Health Futures 2030: Growing up in a Digital World Commission, on the convergence of digital health, AI and universal health coverage, which will consult between October 2019 and December 2021 (*97*).

4.6 Bioethics laws and policies

Bioethics laws and policies play a role in regulating the use of AI, and several bioethics laws have been revised in recent years to include recognition of the growing use of AI in science, health care and medicine. The French Government's most recent revision of its national bioethics law (98), which was endorsed in 2019, establishes standards to address the rapid growth of digital technologies in the health-care system. It includes standards for human supervision, or human warranty, that require evaluation by patients and clinicians at critical points in the development and deployment of AI. It also supports free, informed consent for the use of data and the creation of a secure national platform for the collection and processing of health data.

4.7 Regulatory considerations

Regulation of AI technologies is likely to be developed and implemented by health regulatory authorities responsible for ensuring the safety, efficacy and appropriate use of technologies for health care and therapeutic development. A WHO expert group that is preparing considerations for the regulation of AI for health has discussed areas that should be considered by stakeholders, including developers and regulators, in examining new AI technologies. They include documentation and transparency, risk management and the life-cycle approach, data quality, analytical and clinical validation, engagement and collaboration, and privacy and data protection. Many regulatory authorities are preparing considerations and frameworks for the use of AI, and they should be examined, potentially with the relevant regulatory agency. Governance of AI through regulatory frameworks and the ethical principles that should be considered are discussed in section 9.5.

5. KEY ETHICAL PRINCIPLES FOR USE OF ARTIFICIAL INTELLIGENCE FOR HEALTH

Ethical principles for the application of AI for health and other domains are intended to guide developers, users and regulators in improving and overseeing the design and use of such technologies. Human dignity and the inherent worth of humans are the central values upon which all other ethical principles rest.

An ethical principle is a statement of a duty or a responsibility in the context of the development, deployment and continuing assessment of AI technologies for health. The ethical principles described below are grounded in basic ethical requirements that apply to all persons and that are considered noncontroversial. The requirements are as follows.

- Avoid harming others (sometimes called "Do no harm" or nonmaleficence).
- Promote the well-being of others when possible (sometimes called "beneficence"). Risks of harm should be minimized, while maximizing benefits. Expected risks should be balanced against expected benefits.
- Ensure that all persons are treated fairly, which includes the requirement to ensure that no person or group is subject to discrimination, neglect, manipulation, domination or abuse (sometimes called "justice" or "fairness").
- Deal with persons in ways that respect their interests in making decisions about their lives and their person, including health-care decisions, according to informed understanding of the nature of the choice to be made, its significance, the person's interests and the likely consequences of the alternatives (sometimes called "respect for persons" or "autonomy").

Additional moral requirements can be derived from this list of fundamental moral requirements. For example, safeguarding and protecting individual privacy is not only recognized as a legal requirement in many countries but is also important to enable people to control sensitive information about themselves and self-determination (respect for their autonomy) and to avoid harm.

These ethical principles are intended to provide guidance to stakeholders about how basic moral requirements should direct or constrain their decisions and actions in the specific context of developing, deploying and assessing the performance of AI technologies for health. These principles are also intended to emphasize issues that arise from the use of a technology that could alter relations of moral significance. For example, it has long been recognized that health-care providers have a special duty to advance these values with respect to patients because of the centrality of health to

individual well-being, because of the dependence of patients on health professionals for information about their diagnosis, prognosis and the relative merits of the available treatment or prevention options, and the importance of free and open exchange of information to the provider–patient relationship. If AI systems are used by health-care workers to conduct clinical tasks or to delegate clinical tasks that were once reserved for humans, programmers who design and program such AI technologies should also adhere to these ethical obligations.

Thus, the ethical principles are important for all stakeholders who seek guidance in the responsible development, deployment and evaluation of AI technologies for health, including clinicians, systems developers, health system administrators, policy-makers in health authorities, and local and national governments. The ethical principles listed here should encourage and assist governments and public sector agencies to keep pace with the rapid evolution of AI technologies through legislation and regulation and should empower medical professionals to use AI technologies appropriately.
Ethical principles should also be embedded within professional and technological standards for AI. Software engineers already are guided by standards such as for fitness for purpose, documentation and provenance, and version control. Standards are required to guide the interoperability and design of a program, for continuing education of those who develop and use such technologies and for governance. Moreover, the standards for the evaluation and external audit of systems are evolving in the context of their use. In health computing, there are standards for system integration, electronic health records, system interoperability, implementation and programming structures.

Although ethical principles do not always clearly address limitations in the uses of such technologies, governments should ban or restrict the use of AI or other technologies if they violate or imperil the exercise of human rights, do not conform to other principles or regulations or would be introduced in unprepared or other inappropriate contexts. For example, many countries lack data protection laws or have inadequate regulatory frameworks to guide the introduction of AI technologies.

The claim that certain basic moral requirements must constrain and guide the conduct of persons can also be expressed in the language of human rights. Human rights are intended to capture a basic set of moral and legal requirements for conduct to which every person is entitled regardless of race, sex, nationality, ethnicity, language, religion or any other feature. These rights include human dignity, equality, non-discrimination, privacy, freedom, participation, solidarity and accountability.

Machine-learning systems could advance the protection and enforcement of human rights (including the human right to health) but could undermine core human rights such as non-discrimination and privacy. Human rights and ethical principles are intimately interlinked; because human rights are legally binding, they provide a

powerful framework by which governments, international organizations and private actors are obligated to abide. Private sector actors have the responsibility to respect human rights, independently of state obligations. In fulfilling this responsibility, private sector actors must take continuous proactive and reactive steps to ensure that they do not abuse or contribute to the abuse of human rights.

The existence of a human rights framework does not, however, obviate the need for continuing ethical deliberation. Indeed, much of ethics is intended to expand upon and complement the norms and obligations established in human rights agreements. In many situations, multiple ethical considerations are relevant and require weighing up and balancing to accommodate the multiple principles at stake. An ethically acceptable decision depends on consideration of the full range of appropriate ethical considerations, ensuring that multiple perspectives are factored into the analysis and creating a decision-making process that stakeholders will consider fair and legitimate.

This guidance identifies six ethical principles to guide the development and use of AI technology for health. While ethical principles are universal, their implementation may differ according to the cultural, religious and other social context. Many of the ethical issues arising in the use of AI and machine learning are not completely new but have arisen for other applications of information and communication technologies for health, such as use of any computer to track a disease or make a diagnosis or prognosis. Computers were performing these tasks with various programs long before AI became noteworthy. Ethical guidance and related principles have been articulated for fields such as telemedicine and data-sharing. Likewise, several ethical frameworks have been developed for AI in general, outside the health sector. (See section 4.) The ethical principles listed here are those identified by the WHO Expert Group as the most appropriate for the use of AI for health.

5.1 Protect autonomy

Adoption of AI can lead to situations in which decision-making could be or is in fact transferred to machines. The principle of autonomy requires that any extension of machine autonomy not undermine human autonomy. In the context of health care, this means that humans should remain in full control of health-care systems and medical decisions. AI systems should be designed demonstrably and systematically to conform to the principles and human rights with which they cohere; more specifically, they should be designed to assist humans, whether they be medical providers or patients, in making informed decisions. Human oversight may depend on the risks associated with an AI system but should always be meaningful and should thus include effective, transparent monitoring of human values and moral considerations. In practice, this could include deciding whether to use an AI system for a particular health-care decision, to vary the level of human discretion and decision-making and to develop AI technologies that can rank decisions when appropriate (as opposed to a single decision). These practices

can ensure a clinician can override decisions made by AI systems and that machine autonomy can be restricted and made "intrinsically reversible".

Respect for autonomy also entails the related duties to protect privacy and confidentiality and to ensure informed, valid consent by adopting appropriate legal frameworks for data protection. These should be fully supported and enforced by governments and respected by companies and their system designers, programmers, database creators and others. AI technologies should not be used for experimentation or manipulation of humans in a health-care system without valid informed consent. The use of machine-learning algorithms in diagnosis, prognosis and treatment plans should be incorporated into the process for informed and valid consent. Essential services should not be circumscribed or denied if an individual withholds consent and that additional incentives or inducements should not be offered by either a government or private parties to individuals who do provide consent.

Data protection laws are one means of safeguarding individual rights and place obligations on data controllers and data processors. Such laws are necessary to protect privacy and the confidentiality of patient data and to establish patients' control over their data. Construed broadly, data protection laws should also make it easy for people to access their own health data and to move or share those data as they like. Because machine learning requires large amounts of data – big data – these laws are increasingly important.

5.2 Promote human well-being, human safety and the public interest

AI technologies should not harm people. They should satisfy regulatory requirements for safety, accuracy and efficacy before deployment, and measures should be in place to ensure quality control and quality improvement. Thus, funders, developers and users have a continuous duty to measure and monitor the performance of AI algorithms to ensure that AI technologies work as designed and to assess whether they have any detrimental impact on individual patients or groups.

Preventing harm requires that use of AI technologies does not result in any mental or physical harm. AI technologies that provide a diagnosis or warning that an individual cannot address because of lack of appropriate, accessible or affordable health care should be carefully managed and balanced against any "duty to warn" that might arise from incidental and other findings, and appropriate safeguards should be in place to protect individuals from stigmatization or discrimination due to their health status.

5.3 Ensure transparency, explainability and intelligibility

AI should be intelligible or understandable to developers, users and regulators. Two broad approaches to ensuring intelligibility are improving the transparency and explainability of AI technology.

Transparency requires that sufficient information (described below) be published or documented before the design and deployment of an AI technology. Such information should facilitate meaningful public consultation and debate on how the AI technology is designed and how it should be used. Such information should continue to be published and documented regularly and in a timely manner after an AI technology is approved for use.

Transparency will improve system quality and protect patient and public health safety. For instance, system evaluators require transparency in order to identify errors, and government regulators rely on transparency to conduct proper, effective oversight. It must be possible to audit an AI technology, including if something goes wrong. Transparency should include accurate information about the assumptions and limitations of the technology, operating protocols, the properties of the data (including methods of data collection, processing and labelling) and development of the algorithmic model.

AI technologies should be explainable to the extent possible and according to the capacity of those to whom the explanation is directed. Data protection laws already create specific obligations of explainability for automated decision-making. Those who might request or require an explanation should be well informed, and the educational information must be tailored to each population, including, for example, marginalized populations. Many AI technologies are complex, and the complexity might frustrate both the explainer and the person receiving the explanation. There is a possible trade-off between full explainability of an algorithm (at the cost of accuracy) and improved accuracy (at the cost of explainability).

All algorithms should be tested rigorously in the settings in which the technology will be used in order to ensure that it meets standards of safety and efficacy. The examination and validation should include the assumptions, operational protocols, data properties and output decisions of the AI technology. Tests and evaluations should be regular, transparent and of sufficient breadth to cover differences in the performance of the algorithm according to race, ethnicity, gender, age and other relevant human characteristics. There should be robust, independent oversight of such tests and evaluation to ensure that they are conducted safely and effectively.

Health-care institutions, health systems and public health agencies should regularly publish information about how decisions have been made for adoption of an AI technology and how the technology will be evaluated periodically, its uses, its known limitations and the role of decision-making, which can facilitate external auditing and oversight.

5.4 Foster responsibility and accountability

Humans require clear, transparent specification of the tasks that systems can perform and the conditions under which they can achieve the desired level of performance; this helps to ensure that health-care providers can use an AI technology responsibly. Although AI technologies perform specific tasks, it is the responsibility of human stakeholders to ensure that they can perform those tasks and that they are used under appropriate conditions.

Responsibility can be assured by application of "human warranty", which implies evaluation by patients and clinicians in the development and deployment of AI technologies. In human warranty, regulatory principles are applied upstream and downstream of the algorithm by establishing points of human supervision. The critical points of supervision are identified by discussions among professionals, patients and designers. The goal is to ensure that the algorithm remains on a machine-learning development path that is medically effective, can be interrogated and is ethically responsible; it involves active partnership with patients and the public, such as meaningful public consultation and debate (*101*). Ultimately, such work should be validated by regulatory agencies or other supervisory authorities.

When something does go wrong in application of an AI technology, there should be accountability. Appropriate mechanisms should be adopted to ensure questioning by and redress for individuals and groups adversely affected by algorithmically informed decisions. This should include access to prompt, effective remedies and redress from governments and companies that deploy AI technologies for health care. Redress should include compensation, rehabilitation, restitution, sanctions where necessary and a guarantee of non-repetition.

The use of AI technologies in medicine requires attribution of responsibility within complex systems in which responsibility is distributed among numerous agents. When medical decisions by AI technologies harm individuals, responsibility and accountability processes should clearly identify the relative roles of manufacturers and clinical users in the harm. This is an evolving challenge and remains unsettled in the laws of most countries. Institutions have not only legal liability but also a duty to assume responsibility for decisions made by the algorithms they use, even if it is not feasible to explain in detail how the algorithms produce their results.

To avoid diffusion of responsibility, in which "everybody's problem becomes nobody's responsibility", a faultless responsibility model ("collective responsibility"), in which all the agents involved in the development and deployment of an AI technology are held responsible, can encourage all actors to act with integrity and minimize harm. In such a model, the actual intentions of each agent (or actor) or their ability to control an outcome are not considered.

5.5 Ensure inclusiveness and equity

Inclusiveness requires that AI used in health care is designed to encourage the widest possible appropriate, equitable use and access, irrespective of age, gender, income, ability or other characteristics. Institutions (e.g. companies, regulatory agencies, health systems) should hire employees from diverse backgrounds, cultures and disciplines to develop, monitor and deploy AI. AI technologies should be designed by and evaluated with the active participation of those who are required to use the system or will be affected by it, including providers and patients, and such participants should be sufficiently diverse. Participation can also be improved by adopting open-source software or making source codes publicly available.

AI technology – like any other technology – should be shared as widely as possible. AI technologies should be available not only in HIC and for use in contexts and for needs that apply to high-income settings but they should also be adaptable to the types of devices, telecommunications infrastructure and data transfer capacity in LMIC. AI developers and vendors should also consider the diversity of languages, ability and forms of communication around the world to avoid barriers to use. Industry and governments should strive to ensure that the "digital divide" within and between countries is not widened and ensure equitable access to novel AI technologies.
AI technologies should not be biased. Bias is a threat to inclusiveness and equity because it represents a departure, often arbitrary, from equal treatment. For example, a system designed to diagnose cancerous skin lesions that is trained with data on one skin colour may not generate accurate results for patients with a different skin colour, increasing the risk to their health.

Unintended biases that may emerge with AI should be avoided or identified and mitigated. AI developers should be aware of the possible biases in their design, implementation and use and the potential harm that biases can cause to individuals and society. These parties also have a duty to address potential bias and avoid introducing or exacerbating health-care disparities, including when testing or deploying new AI technologies in vulnerable populations.

AI developers should ensure that AI data, and especially training data, do not include sampling bias and are therefore accurate, complete and diverse. If a particular racial or ethnic minority (or other group) is underrepresented in a dataset, oversampling of that group relative to its population size may be necessary to ensure that an AI technology achieves the same quality of results in that population as in better-represented groups.

AI technologies should minimize inevitable power disparities between providers and patients or between companies that create and deploy AI technologies and those that use or rely on them. Public sector agencies should have control over the data collected

by private health-care providers, and their shared responsibilities should be defined and respected. Everyone – patients, health-care providers and health-care systems – should be able to benefit from an AI technology and not just the technology providers. AI technologies should be accompanied by means to provide patients with knowledge and skills to better understand their health status and to communicate effectively with health-care providers. Future health literacy should include an element of information technology literacy.

The effects of use of AI technologies must be monitored and evaluated, including disproportionate effects on specific groups of people when they mirror or exacerbate existing forms of bias and discrimination. Special provision should be made to protect the rights and welfare of vulnerable persons, with mechanisms for redress if such bias and discrimination emerges or is alleged.

5.6 Promote artificial intelligence that is responsive and sustainable

Responsiveness requires that designers, developers and users continuously, systematically and transparently examine an AI technology to determine whether it is responding adequately, appropriately and according to communicated expectations and requirements in the context in which it is used. Thus, identification of a health need requires that institutions and governments respond to that need and its context with appropriate technologies with the aim of achieving the public interest in health protection and promotion. When an AI technology is ineffective or engenders dissatisfaction, the duty to be responsive requires an institutional process to resolve the problem, which may include terminating use of the technology.

Responsiveness also requires that AI technologies be consistent with wider efforts to promote health systems and environmental and workplace sustainability. AI technologies should be introduced only if they can be fully integrated and sustained in the health-care system. Too often, especially in under-resourced health systems, new technologies are not used or are not repaired or updated, thereby wasting scare resources that could have been invested in proven interventions. Furthermore, AI systems should be designed to minimize their ecological footprints and increase energy efficiency, so that use of AI is consistent with society's efforts to reduce the impact of human beings on the earth's environment, ecosystems and climate. Sustainability also requires governments and companies to address anticipated disruptions to the workplace, including training of health-care workers to adapt to use of AI and potential job losses due to the use of automated systems for routine health-care functions and administrative tasks.

6. ETHICAL CHALLENGES TO USE OF ARTIFICIAL INTELLIGENCE FOR HEALTH CARE

Several ethical challenges are emerging with the use of AI for health, many of which are especially relevant to LMIC. These challenges must be addressed if AI technologies are to support achievement of universal health coverage. Use of AI to extend health-care coverage and services in marginalized communities in HIC can raise similar ethical concerns, including an enduring digital divide, lack of good-quality data, collection of data that incorporate clinical biases (as well as inappropriate data collection practices) and lack of treatment options after diagnosis.

6.1 Assessing whether artificial intelligence should be used

There are risks of overstatement of what AI can accomplish, unrealistic estimates of what could be achieved as AI evolves and uptake of unproven products and services that have not been subjected to rigorous evaluation for safety and efficacy (93). This is due partly to the enduring appeal of "technological solutionism", in which technologies such as AI are used as a "magic bullet" to remove deeper social, structural, economic and institutional barriers (*102*). The appeal of technological solutions and the promise of technology can lead to overestimation of the benefits and dismissal of the challenges and problems that new technologies such as AI may introduce. This can result in an unbalanced health-care policy and misguided investments by countries that have few resources and by HIC that are under pressure to reduce public expenditure on health care (*103*). It can also divert attention and resources from proven but underfunded interventions that would reduce morbidity and mortality in LMIC.

First, the AI technology itself may not meet the standards of scientific validity and accuracy that are currently applied to medical technologies. For example, digital technologies developed in the early stages of the COVID-19 pandemic did not necessarily meet any objective standard of efficacy to justify their use (*104*). AI technologies have been introduced as part of the pandemic response without adequate evidence, such as from randomized clinical trials, or safeguards (*9*). An emergency does not justify deployment of unproven technologies (*104*); in fact, efforts to ensure that resources were allocated where they were most urgently needed should have heightened the vigilance of both companies and governments (such as regulators and ministries of health) to ensure that the technologies were accurate and effective.

Secondly, the benefits of AI may be overestimated when erroneous or overly optimistic assumptions are made about the infrastructure and institutional context in which the technologies will be used and where the intrinsic requirements for use of the technology cannot be met. In some low-income countries, financial resources and

information and communication technology infrastructure lag those of HIC, and the significant investments that would be required might discourage use. This is discussed in greater detail in section 6.2. The quality and availability of data may not be adequate for use of AI, especially in LMIC. There is a danger that poor-quality data will be collected for AI training, which may result in models that predict artefacts in the data instead of actual clinical outcomes. There may also be no data, which, with poor-quality data, could distort the performance of an algorithm, resulting in inaccurate performance, or an AI technology might not be available for a specific population because of insufficient usable data. Additionally, significant investment may be required to make non-uniform data sets collected in LMIC usable. Compilation of data in resource-poor settings is difficult and time-consuming, and the additional burden on community health workers should be considered. Data are unlikely to be available on the most vulnerable or marginalized populations, including those for whom health-care services are lacking, or they might be inaccurate. Data may also be difficult to collect because of language barriers, and mistrust may lead people to provide incorrect or incomplete information. Often, irrelevant data are collected, which can undermine the overall quality of a dataset.[4] Broader concern about the collection and use of data, as well as bias in data, is discussed below.

There may not be appropriate or enforceable regulations, stakeholder participation or oversight, all of which are required to ensure that ethical and legal concerns can be addressed and human rights are not violated. For example, AI technologies may be introduced in countries without up-to-date data protection and confidentiality laws (especially for health-related data) or without the oversight of data protection authorities to rigorously protect confidentiality and the privacy of individuals and communities. Furthermore, regulatory agencies in LMIC may not have the capacity or expertise to assess AI technologies to ensure that systematic errors do not affect diagnosis, surveillance and treatment.

Thirdly, there may be enough ethical concern about a use case or a specific AI technology, even if it provides accurate, useful information and insights, to discourage a particular use. An AI technology that can predict which individuals are likely to develop type 2 diabetes or HIV infection could provide benefits to an at-risk individual or community but could also give rise to unnecessary stigmatization of individuals or communities, whose choices and behaviour are questioned or even criminalized, result in over-medicalization of otherwise healthy individuals, create unnecessary stress and anxiety and expose individuals to aggressive marketing by pharmaceutical companies and other for-profit health-care services (*105*). Furthermore, certain AI technologies, if not deployed carefully, could exacerbate disparities in health care, including those related to ethnicity, socioeconomic status or gender.

[4] Presentation by Dr Amel Ghoulia, Bill & Melinda Gates Foundation, 3 October 2019, to the WHO working group on ethics and governance of AI for health.

Fourthly, like all new heath technologies, even if an AI technology does not trigger an ethics warning, its benefits may not be justified by the extra expense or cost (beyond information and communication technology infrastructure) associated with the procurement, training and technology investment required (*43*). Robotic surgery may produce better outcomes, but the opportunity costs associated with the investment must also be considered.

Fifthly, enough consideration may not be given to whether an AI technology is appropriate and adapted to the context of LMIC, such as diverse languages and scripts in a country or among countries (*9*). Lack of investment in, for example, translation can mean that certain applications do not operate correctly or simply cannot be used by a population. Such lack of foresight points to a wider problem, which is that many AI technologies are designed by and for high-income populations and by individuals or companies with inadequate understanding of the characteristics of the target populations in LMIC.

Unrealistic expectations of what AI can achieve may, however, unnecessarily discourage its use. Thus, machines and algorithms (and the data used for algorithms) are expected in the public imagination to be perfect, while humans can make mistakes. Medical professionals might overestimate their ability to perform tasks and ignore or underestimate the value of algorithmic decision tools, for which the challenges can be managed and for which evidence indicates a measurable benefit. Not using the technology could result in avoidable morbidity and mortality, making it blameworthy not to use a certain AI technology, especially if the standard of care is already shifting to its use (*106*). For medical professionals to make such an assessment, they require greater transparency with regard to the performance and utility of AI technologies, a principle enumerated in section 5 of this report, as well as effective regulatory oversight. The role of regulatory agencies in ensuring rigorous testing, transparent communication of outcomes and monitoring of performance is discussed in section 9.5.

Even after an AI technology has been introduced into a health-care system, its impact should be evaluated continuously during its real-world use, as should the performance of an algorithm if it learns from data that are different from its training data. Impact assessments can also guide a decision on use of AI in an area of health before and after its introduction (*106*). (See section 7.3.) Assessment of whether to introduce an AI technology in a low-income country or resource-poor setting may lead to a different conclusion from such an assessment in a high-income setting. Risk–benefit calculations that do not favour a specific use of AI in HIC may be interpreted differently for a low-income country that lacks, for example, enough health-care workers to perform certain tasks or which would otherwise forego use of more accurate diagnostic instruments, such that individuals receive inaccurate diagnoses and the wrong treatment.

The use of AI to resource-poor contexts should, however, be extended carefully to avoid situations in which large numbers of people receive accurate diagnoses of a health condition but have no access to appropriate treatment. Health-care workers have a duty to provide treatment after testing for and confirmation of disease, and the relatively low cost at which AI diagnostics can be deployed should be accompanied by careful planning to ensure that people are not left without treatment.[5] Prediction tools for anticipating a disease outbreak will have to be complemented by robust surveillance systems and other effective measures.

6.2 Artificial intelligence and the digital divide

Many LMIC have sophisticated economies and digital infrastructure, while others, such as India, have both world-class digital infrastructure and millions of people without electricity. The countries with the greatest challenges to adoption of AI are classified as least developed; however, AI could allow those countries to leapfrog existing models of health-care delivery to improve health outcomes (*23*).

One challenge that could affect the uptake of AI is the "digital divide", which refers to uneven distribution of access to, use of or effect of information and communication technologies among any number of distinct groups. Although the cost of digital technologies is falling, access has not become more equitable. For example, 1.2 billion women (327 million fewer women than men) in LMIC do not use mobile Internet services because they cannot afford to or do not trust the technology, even though the cost of the devices should continue to fall (*107*). Gender is only one dimension of the digital divide; others are geography, culture, religion, language and generation. The digital divide begets other disparities and challenges, many of which affect the use of AI, and AI itself can reinforce and exacerbate the disparity. Thus, in 2019, the United Nations Secretary-General's High-level Panel on Digital Cooperation (*108*) recommended that

> by 2030, every adult should have affordable access to digital networks, as well as digitally enabled financial and health services, as a means to make a substantial contribution to achieving the Sustainable Development Goals.

The human and technical resources required to realize the benefits of digital technologies fully are also unequally distributed, and infrastructure to operate digital technologies may be limited or inexistent. Some technologies require an electricity grid and information and communication technology infrastructure, including electrification, Internet connectivity, wireless and mobile networks and devices. Solar energy may provide a path forward for many countries if the climate is appropriate, as investment is increasing and the cost of solar energy has decreased dramatically in the past decade (*109*). Nevertheless, at present, an estimated 860 million people

[5] The International Council of Nurses noted: "'Ethical issues may arise if there is the capability of AI diagnostics but not the capacity to provide treatment. Issues like this have arisen in the field of endoscopy in some countries where some diagnostic services for screening are withheld because of the limited access to surgical services." Communication from the International Council of Nurses to WHO on 6 January 2021.

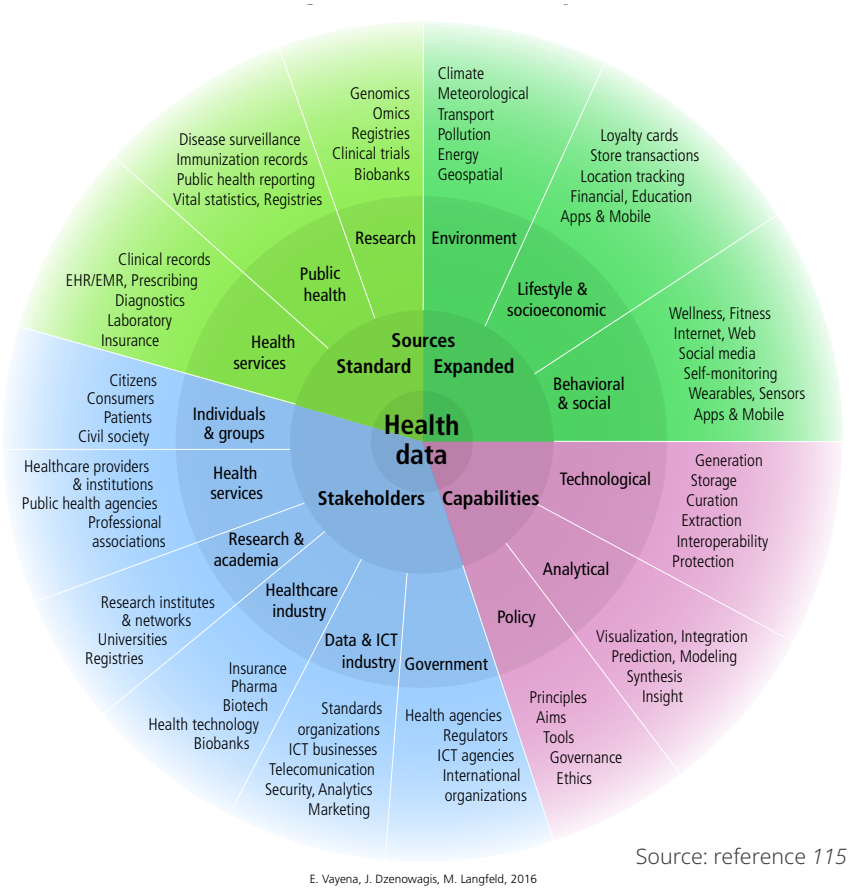
worldwide do not have access to electricity, including 600 million people in sub-Saharan Africa, and there is growing pressure on the electrical grid in cities due to urbanization (*110*). Even in high-income economies with near-universal electrification and enough resources, the digital divide has persisted. In the USA, for example, millions of people in rural areas and in cities still lack access to high-speed broadband services, and 60% of health-care facilities outside metropolitan areas also lack broadband (*111*).

Even as countries overcome the digital divide, technology providers should be required to provide infrastructure, services and programs that are interoperable, so that different platforms and applications can work seamlessly with one another, as well as affordable devices (for example, smartphones) that do not require consumers to trade privacy for affordability (*112*). This will ensure that the emerging digital health-care system is not fragmented and is equitable.

6.3 Data collection and use

The collection, analysis and use of health data, including from clinical trials, laboratory results and medical records, is the bedrock of medical research and the practice of medicine. Over the past two decades, the data that qualify as health data have expanded dramatically. They now include massive quantities of personal data about individuals from many sources, including genomic data, radiological images, medical records and non-health data converted into health data (*113*). The various types of data, collectively known as "biomedical big data", form a health data ecosystem that includes data from standard sources (e.g. health services, public health, research) and further sources (environmental, lifestyle, socioeconomic, behavioural and social) (Fig. 1) (*114*).

Fig. 1. Health data ecosystem



Source: reference *115*

Thus, there are many more sources of health data, entities

that wish to make use of such data and commercial and non-commercial applications. The development of a successful AI system for use in health care relies on high-quality data for both training the algorithm and validating the algorithmic model.

The potential benefits of biomedical big data can be ethically important, as AI technologies based on high-quality data can improve the speed and accuracy of diagnosis, improve the quality of care and reduce subjective decision-making. The ubiquity of health data and the potential sensitivity of health care to data indicate possible benefits. Health care is still lagging in the adoption of data science and AI as compared with other sectors (although some would disagree), and individuals informed of the potential benefits of the collection and use of such data might support use of such data for their personal benefit or that of a wider group.[6]

Several concerns may undermine effective use of health data in AI-guided research and drug development. Concern about the use of health data is not limited to their use in AI, although AI has exacerbated the problem. One concern with health data is their quality, especially with those from LMIC (see above). Furthermore, training data will always have one or more systemic biases because of under-representation of a gender, age, race, sexual orientation or other characteristic. These biases will emerge during modelling and subsequently diffuse through the resulting algorithm (*103*). Concern about the impact of bias is discussed in section 6.6.

A second major concern is safeguarding individual privacy. The collection, use, analysis and sharing of health data have consistently raised broad concern about individual privacy, because lack of privacy may either harm an individual (such as future discrimination on the basis of one's health status) or cause a wrong, such as affecting a person's dignity if sensitive health data are shared or broadcast to others (*116*). There is a risk that sharing or transferring data leaves them vulnerable to cyber-theft or accidental disclosure (*116*). Recommendations generated by an algorithm from an individual's health data also raise privacy concerns, as a person may expect that such "new" health data are private (*116*), and it may be illegal for third parties to use "new" health data. Such privacy concerns are heightened for stigmatized and vulnerable populations, for whom data disclosure can lead to discrimination or punitive measures (*117*). There is also concern about the rights of children (*118*), which could include future discrimination based on the data accumulated about a child, children's ability to protect their privacy and their autonomy to make choices about their health care. Measures to collect data or track an individual's status and to construct digital identities to store such information have accelerated during the COVID-19 pandemic. See Box 2.

[6] Presentation by Dr Andrew Morris, Health Data Research United Kingdom, 3 October 2019 to the WHO working group on ethics and governance of AI for health.

Box 2. The emergence of digital identification in the COVID-19 pandemic

The COVID-19 pandemic is expanding and accelerating the creation of infrastructure for digital identities to store health data for several uses. In China, a QR code system has been established from the digital payment system established by Alipay, a mobile and online payment platform, to introduce an "Alipay Health Code", in which the data collected are used to establish an algorithm to "draw automated conclusions as to whether someone is a contagion risk" (*119*). For a national programme to vaccinate millions of people against SARS-Cov2, India may use its national digital ID system, Aadhar, to avoid duplication and to track beneficiaries (*120*). Many entities around the world, including travel firms, airports, some governments and political leaders, as well as the digital ID industry, are calling for the introduction of immunity passports or a digital "credential given to a person who is assumed to be immune from SARS-CoV2 and so protected against re-infection" (*121*). In some countries, technologies such as proximity-tracking applications have been credited with improving the response to the pandemic, because there was already a system in place to support the use of such technologies, effective communication, widespread adoption and a "social compact" between policy-makers and the public (*122*).

For many of these technologies, however, there is concern about whether they are effective (scientifically valid), whether they will create forms of discrimination or targeting of certain populations and whether they may exclude certain segments of the population or not be applicable by people who do not have access to the appropriate technology and infrastructure. They also raise concern about the generation of a permanent digital identity for individuals linked to their health and personal data, for which they may not have given consent, which could permanently undermine individual autonomy and privacy (*123*). In particular, there is concern that governments could use such information to establish mass surveillance or scoring systems to monitor everyday activities, or companies could use such data and systems for other purposes (*124*).

A third major concern is that health data collected by technology providers may exceed what is required and that such excess data, so-called "behavioural data surplus" (*125*), is repurposed for uses that raise serious ethical, legal and human rights concerns. The uses might include sharing such data with government agencies so that they can exercise control or use punitive measures against individuals (*104*). Such repurposing, or "function creep", is a challenge that predates but is heightened by the use of AI for health care. For example, in early 2021, the Singapore Government admitted that data obtained from its COVID-19 proximity-tracing application (Trace Together) could also be accessed "for the purpose of criminal investigation", despite prior assurances that this would not be permitted (*126*). In February 2021, legislation was introduced to restrict the use of such data for only the most "serious" criminal investigations, such as for murder or terrorism-related charges, with penalties for any unauthorized use (*127*).

Such data may also be shared with companies that use them to develop an AI technology for marketing goods and services or to create prediction-based products to be used, for example, by an insurance firm (*128*) or a large technology company. Such uses of health data, often unknown to those who have supplied the data, have generated front-page headlines and public concern (*129*). The provision of health data to commercial entities has also resulted in the filing of legal actions by individuals whose health data (de-identified) have been disclosed on behalf of all affected individuals. See Box 3.

Box 3. Dinerstein vs Google

Google announced a strategic partnership with the University of Chicago and the University of Chicago Medicine in the USA in May 2017 (*130*). The aim of the partnership was to develop novel machine-learning tools to predict medical events such as unexpected hospital admissions. To realize this goal, the University shared hundreds of thousands of "de-identified" patients' records with Google. One of the University's patients, Matt Dinerstein, filed a class action complaint against the University and Google in June 2019 on behalf of all patients whose records were disclosed (*131*).

Dinerstein brought several claims, including breach of contract, against the University and Google, alleging prima facie violation of the US Health Insurance Portability and Accountability Act. According to an article published in 2018 by the defendants (*132*), the patients' medical records shared with Google "were de-identified, except that dates of service were maintained in the (...) dataset". The dataset also included "free-text medical notes" (*132*). Dinerstein accused the defendants of insufficient anonymization of the records, putting the patients' privacy at risk. He alleged that the patients could easily be re-identified by Google by combining the records with other available data sets, such as geolocation data from Google Maps (by so-called "data triangulation"). Moreover, Dinerstein asserted that the University had not obtained express consent from each patient to share their medical records with Google, despite the technology giant's commercial interest in the data.

The issue of re-identification was largely avoided by the district judge, who dismissed Dinerstein's lawsuit in September 2020. The reasons given for dismissal included Dinerstein's failure to demonstrate damages that had occurred because of the partnership. This case illustrates the challenges of lawsuits related to data-sharing and highlights the lack of adequate protection of the privacy of health data. In the absence of ethical guidelines and adequate legislation, patients may have difficulty in maintaining control of their personal medical information, particularly in circumstances in which the data can be shared with third parties and in the absence of safeguards against re-identification.

This case study was written by Marcelo Corrales Compagnucci (CeBIL Copenhagen), Sara Gerke (Harvard Law School) and Timo Minssen (CeBIL Copenhagen).

Some companies have already collected large quantities of health data through their products and services, to which users voluntarily supply health data (user-generated health data) (*133*). They may acquire further data through a data aggregator or broker (*134*) or may rely on governments to aggregate data that can be used by public, not-for-profit and private sector entities (*135*). Such data may include "mundane" data that were not originally characterized as "health data"; however, machine learning can elicit sensitive details from such ordinary personal data and thus transform them into a special category of sensitive data (*136*) that may require protection.

Concern about the commercialization of health data includes individual loss of autonomy, a principle stated in section 5, loss of control over the data (with no explicit consent to such secondary use), how such data (or outcomes generated by such data) may be used by the company or a third party, with concern that companies are allowed to profit from the use of such data, and concern about privacy, as companies may not meet the duty of confidentiality, whether purposefully or inadvertently (for example due to a data breach) (*137*). Thus, once an individual's medical history is exposed, it cannot be replaced in the same way as a new credit card can be obtained after a breach.

Data colonialism

A fourth concern with biomedical big data is that it may foster a divide between those who accumulate, acquire, analyse and control such data and those who provide the data but have little control over their use. This is especially true with respect to data collected from underrepresented groups, many of which are predominantly in LMIC, often with the broad ambition of collecting data for development or for humanitarian ends rather than to promote local economic development and governance (*138*). Insufficient data from underrepresented groups affect them negatively, and attention has focused on either encouraging such groups to provide data or instituting measures to collect data. Generating more data from LMIC, however, also carries risks, including "data colonialism", in which the data are used for commercial or non-commercial purposes without due respect for consent, privacy or autonomy. Collection of data without the informed consent of individuals for the intended uses (commercial or otherwise) undermines the agency, dignity and human rights of those individuals; however, even informed consent may be insufficient to compensate for the power dissymmetry between the collectors of data and the individuals who are the sources. This is a particular concern because of the possibility that companies in countries with strict regulatory frameworks and data protection laws could extend data collection to LMIC without such control. While regulatory frameworks such as the EU's GDPR include an "extra-territorial" clause that requires compliance with its standards outside the EU, entities are not obliged to provide a right of redress as guaranteed under the EU GDPR, and companies may use such data but not provide appropriate products and services to the underserved communities and countries

from which the data were obtained. Individuals in these regions therefore have little or no knowledge of how their data are being used, by a government or company, no opportunity to provide any form of consent for how the data could be used and often less bargaining power if recommendations based on the data have an adverse effect on an individual or a community (*139*).

Mechanisms for safeguarding privacy – do they work?

When meaningful consent is possible, it can overcome many concerns, including those related to privacy. Yet, true informed consent is increasingly infeasible in an era of biomedical big data, especially in an environment driven mainly by companies seeking to generate profits from the use of data (*113*). The scale and complexity of biomedical big data make it impossible to keep track of and make meaningful decisions about all uses of personal data (*113*). All the potential uses of health data may not be known, as they may eventually be linked to and used for a purpose that is far removed from the original intention. Patients may be unable to consent to current and future uses of their health data, such as for population-level data analytics or predictive-risk modelling (*113*). Even if a use lends itself to consent, the procedures may fall short, individuals might not be able to consent, such as because they have insufficient access to a health data system, or access to health care is perceived or actually denied if consent is not provided.

One concern is in the management of use of health data (probably collected for different purposes and not necessarily to support the use of AI) after an individual has died. Such data could provide numerous benefits for medical research (*140*), to improve understanding of the causes of cancer (*141*) or to increase the diversity of data used for medical AI. These data must, however, also be protected against unauthorized use. Existing laws either define limited circumstances in which such data can be used or restrict how they can be used (*142*). In the GDPR, a data protection law does not apply to deceased persons, and, under Article 27, EU Member States "may provide for rules regarding the processing of personal data of deceased persons" (*143*). Proposals have been made to improve the sharing of such data through voluntary and participatory approaches by which individuals can provide broad or selective consent for use of their data after death, much as individuals can provide consent for use of their organs for medical research (*143*).

If patients' privacy cannot be safeguarded by consent mechanisms, other privacy safeguards, including a data holder's duty of confidentiality, also have shortcomings. Although confidentiality is a well-recognized pillar of medical practice, the duty of confidentiality may not be sufficient to cover the many types of data now used to guide AI health technologies and may also not be sufficient to control the production and transfer of health data (*113*).

A proactive approach to preserving privacy is de-identification or anonymization or pseudo-anonymization of health data. De-identification prevents connection of personal identifiers to information. Anonymization of personal data is a subcategory of de-identification whereby both direct and indirect personal identifiers are removed, and technical safeguards are used to ensure zero risk of re-identification, whereas de-identified data can be re-identified by use of a key (*144*). Pseudo-anonymization is defined in Article 5 of the GDPR (*145*) as:

> processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

The use of such techniques could safeguard privacy and encourage data-sharing but also raises several concerns and challenges. In the USA for example, fully de-identified health data can be used for other purposes without consent (*146*). De-identification may not always be successful, as "data triangulation" techniques can be used to reconstruct a de-identified, incomplete dataset by a third party for re-identification of an individual (*147*). It may be impossible completely to de-identify some types of data, such as genome sequences, as relationships to other people whose identity and partial sequence are known can be inferred. Such relationships may allow direct identification of small groups and to narrow down identification to families (*128*, *148*).

Anonymization may not be possible during health data collection. For example, in predictive AI, time-course data must be collected from a single individual at several times, obviating anonymization until data at all time points are collected. Furthermore, while anonymization may minimize the risks of (re-)identification of a person, it can reduce the positive benefits of health data, including re-assembly of fragments of an individual's health data into a comprehensive profile of a patient, which is required for some forms of AI such as predictive algorithms of mortality. Furthermore, anonymization may undermine a person's right to control their own data and how it may be used (*113*). Other techniques could be used to preserve privacy, including differential privacy, synthetic data generation and k-anonymity, which are briefly discussed in section 7.1.

6.4 Accountability and responsibility for decision-making with artificial intelligence

This section addresses the challenges of assigning responsibility and accountability for the use of AI for health care, a guiding principle noted in section 5. Much of the momentum of AI is based on the notion that use of such technologies for diagnosis, care or systems could improve clinical and institutional decision-making for health

care. Clinicians and health-care workers have numerous cognitive biases and commit diagnostic errors. The US National Academy of Sciences found that 5% of US adults who seek health advice receive erroneous diagnoses and that such errors account for 10% of all patient deaths (*149*). At the institutional level, machine learning might reduce inefficiency and errors and ensure more appropriate allocation of resources, if the underlying data are both accurate and representative (*149*).

AI-guided decision-making also introduces several trade-offs and risks. One set of trade-offs is associated with the displacement of human judgement and control and concern about using AI to predict a person's health status or the evolution of disease. This is a major ethical and epistemological challenge to humans as the centre of production of knowledge and also to the system of production of knowledge for medicine. These considerations are addressed in section 6.5.

Governments can violate human rights (and companies can fail to respect human rights), undermine human dignity or cause tangible harm to human health and well-being by using AI-guided technologies. These violations may not be foreseen during development of an AI technology and may emerge only once the technology evolves in real-world use. If proactive measures such as greater transparency and continuous updating of training data do not avoid harm, recourse may be made through civil (and occasionally criminal) liability. The use of liability regimes to address harm caused by AI-guided technologies is addressed in section 8.

Responsibility ensures that individuals and entities are held accountable for any adverse effects of their actions and is necessary to maintain trust and to protect human rights. Certain characteristics of AI technologies, however, affect notions of responsibility (and accountability), including their opacity, reliance on human input, interaction, discretion, scalability, capacity to generate hidden insights and the complexity of the software. One challenge to assigning responsibility is the 'control problem' associated with AI, wherein developers and designers of AI may not be held responsible, as AI-guided systems function independently of their developers and may evolve in ways that the developer could claim were not foreseeable (*150*). This creates a responsibility gap, which could place an undue burden on a victim of harm or on the clinician or health-care worker who uses the technology but was not involved in its development or design (*150*, *151*). Assigning responsibility to the developer might provide an incentive to take all possible steps to minimize harm to the patient. Such expectations are already well established for the producers of other commonly used medical technologies, including drug and vaccine manufacturers, medical device companies and medical equipment makers.

The 'control problem' will become ever more salient with the emergence of automated AI. Technology companies are making large investments in automating

the programming of AI technologies, partly because of the scarcity of AI developers. Automation of AI programming, through programs such as BigML, Google AutoML and Data Robot, might be attractive to public health institutions that wish to use AI but lack the budget to hire AI developers (*152*). While automated AI programming might be more accurate, its use might not be fair, ethical or safe in certain situations. If AI programming is automated, the checks and balances provided by the involvement of a human developer to ensure safety and identify errors would also be automated, and the control problem is abstracted one step further away from the patient.

A second challenge is the "many hands problem" or the "'traceability" of harm, which bedevils health-care decision-making systems (*153*) and other complex systems (*154*) even in the absence of AI. As the development of AI involves contributions from many agents, it is difficult, both legally and morally, to assign responsibility (*150*), which is diffused among all the contributors to the AI-guided technology. Participation of a machine in making decisions may also discourage assignment of responsibility to the humans involved in the design, selection and use of the technology (*150*). Diffusion of responsibility may mean that an individual is not compensated for the harm he or she suffers, the harm itself and its cause are not fully detected, the harm is not addressed and societal trust in such technologies may be diminished if it appears that none of the developers or users of such technologies can be held responsible (*155*).

A third challenge to assigning responsibility is the issuance of ethics guidance by technology companies, separately or jointly (*156*). Such guidance sets out norms and standards to which the companies commit themselves to comply publicly and voluntarily. Many companies have issued such guidance in the absence of authoritative or legally binding international standards. Recognition by technology companies that AI technologies for use in health care and other sectors are of public concern and must be carefully designed and deployed to avoid harm, such as violations of human rights or bodily injury, is welcome. Such guidelines may, however, depending on how they are implemented, be little more than "ethics washing" (*150*). First, the public tends to have little or no role in setting such standards (*157*). Secondly, such guidelines tend to apply to the prospective behaviour of companies for the technologies they design and deploy (role responsibility) and not historic responsibility for any harms for which responsibility should be allocated. This creates a responsibility gap, as it does not address causal responsibility or retrospective harm (*150*). Thirdly, monitoring of whether companies are complying with their own guidance tends to be done internally, with little to no transparency, and without enforcement by institutions or mechanisms empowered to act independently to evaluate whether the commitments are being met (*157*, *158*). Finally, these commitments are not legally enforceable if violated (*158*).

AI provides great power and benefits (including the possibility of profit) to those who design and deploy such systems. Thus, reciprocity should apply – companies that reap direct and indirect benefits from AI-guided technologies should also have to shoulder responsibility for any negative consequences (section 8), especially as it is health-care providers who will bear the immediate brunt of any psychological stress if an AI technology causes harm to a patient. Companies should also allow independent audits and oversight of enforcement of its own ethics standards to ensure that the standards are being met and that corrective action is taken if a problem arises.

Accountability for AI-related errors and harm

Clinicians already use many non-AI technologies in diagnosis and treatment, such as X-rays and computer software. As AI technologies are used to assist or improve clinical decision-making and not to replace it, there may be an argument to initially hold clinicians accountable for any harm that results from their use in health care. In the same way as for non-AI technologies, however, this oversimplifies the reasons for harm and who should be held accountable for such harm. If a clinician makes a mistake in using the technology, he or she may be held accountable if they were trained in its use that otherwise may not have been included in their medical training (*159*). Yet, if there is an error in the algorithm or the data used to train the AI technology, for example, accountability might be better placed with those who developed or tested the AI technology rather than requiring the clinician to judge whether the AI technology is providing useful guidance (*159*).

There are other reasons for not holding clinicians solely accountable for decisions made by AI technologies, several of which apply to assigning accountability for the use of non-AI health technologies. First, clinicians do not exercise control over an AI-guided technology or its recommendations (*151*). Secondly, as AI technologies tend to be opaque and may use "black-box" algorithms, a physician may not understand how an AI system converts data into decisions (*151*). Thirdly, the clinician may not have chosen to use the AI technology but does so because of the preferences of the hospital system or of other external decision-makers.

Furthermore, if physicians were made accountable for harm caused by an AI technology, technology companies and developers could avoid accountability, and human users of the technology would become the scapegoats of all faults arising from its use, with no control over the decisions made by the AI technology (*150*). Furthermore, with the emergence of autonomous systems for driving and warfare, there is growing concern about whether humans can exert "meaningful control" over such technologies or whether the technologies will increasingly make decisions independently of human input. (See section 6.5.)

Clinicians should not, however, be fully exempt from accountability for errors in content, in order to avoid "automation bias" or lack of consideration of whether an automated

technology meets their needs or those of the patient (*159*). In automation bias, a clinician may overlook errors that should have been spotted by human-guided decision-making. While physicians must be able to trust an algorithm, they should not ignore their own expertise and judgement and simply rubber-stamp the recommendation of a machine (*160*). Some AI technology may not issue a single decision but a set of options from which a physician must select. If the physician makes the wrong choice, what should the criteria be for holding the physician accountable?

Assignation of accountability is even more complex when a decision is made to use an AI technology throughout a health-care system, as the developer, the institution and the physician may all have played a role in the medical harm, yet none is fully to blame (*149*). In such situations, accountability may rest not with the provider or the developer of the technology but with the government agency or institution that selected, validated and deployed it.

6.5 Autonomous decision-making

Decision-making has not yet been "fully transferred" from humans to machines in health care. While AI is used only to augment human decision-making in the practice of public health and medicine, epistemic authority has, in some circumstances, been displaced, whereby AI systems (such as with the use of computer simulations) are displacing humans from the centre of knowledge production (*161*, *162*). Furthermore, there are signs of full delegation of routine medical functions to AI. Delegation of clinical judgement introduces concern about whether full delegation is legal, as laws increasingly recognize the right of individuals not to be subject to solely automated decisions when such decisions would have a significant effect. Full delegation also creates a risk of automation bias on the part of the provider, as discussed above. Other concerns could emerge if human judgement is increasingly replaced by machine-guided judgement, and wider ethical concern would arise with loss of human control, especially if prediction-based health care becomes the norm. Yet, as for autonomous cars, it is unlikely that AI in medicine will ever achieve full autonomy. It may achieve only conditional automation or require human back-up (*163*).

Implications of replacing human judgement for clinical care

There are benefits of replacing human judgement and of humans ceding control over certain aspects of clinical care. Humans could make worse decisions that are less fair and more biased compared to machines (concern about bias in the use of AI is discussed below). Use of AI systems to make specific, well-defined decisions may be entirely justified if there is compelling clinical evidence that the system performs the task better than a human. Leaving decisions to humans when machines can perform them more rapidly, more accurately and with greater sensitivity and specificity can mean that some patients suffer avoidable morbidity and mortality without the prospect of some offsetting benefit (*106*).

In some cases, automation of routine, mundane functions, such as recording information, could liberate a medical provider to build or enhance a relationship with a patient while AI-guided machines automate certain aspects of caregiving (*24*). Other mundane functions could be fully assumed by AI, such as automatic adjustment of a hospital ward temperature.

The shift to applying AI technologies for more complex areas of clinical care will, however, present several challenges. One is the likely emergence of "peer disagreement" between two competent experts – an AI machine and a doctor (*149*). In such situations, there is no means of combining the decisions or of reasoning with the algorithm, as it cannot be accessed or engaged to change its mind. There are also no clear rules for determining who is right, and if a patient is left to trust either a technology or a physician, the decision may depend on factors that have no basis in the "expertise" of the machine or the doctor. Choosing one of the two leads to an undesirable outcome. If the doctor ignores the machine, AI has added little value (*149*). If the doctor accepts the machine's decision, it may undermine his or her authority and weaken their accountability. Some may argue that the recommendation of an algorithm should be preferred, as it combines the expertise of multiple experts and many data points (*149*).

The challenge of human–computer interactions has been addressed by validating systems, providing appropriate education for users and validating the systems continuously. It may, however, be ethically challenging for doctors to rely on the judgement of AI, as they have to accept decisions based on black-box algorithms (*159*). The widely held convention is that many algorithms, e.g. those based on artificial neural networks or other complex models, are black boxes that make inferences and decisions that are not understood even by their developers (*164*). It may therefore be questioned whether doctors can be asked to act on decisions made by such black-box algorithms. AI should therefore be transparent and explainable, which is listed as a core guiding principle in section 5. Some argue that, if a trade-off must be made between even greater transparency (and explainability) and accuracy, transparency should be preferred. This requirement, however, goes beyond what may be possible or even desirable in a medical context. While it is often possible to explain to a patient why a specific treatment is the best option for a specific condition, it is not always possible to explain how that treatment works or its mechanism of action, because some medical interventions are used before their mode of action is understood (*165*). It may be more important to explain how a system has been validated and whether a particular use falls within the parameters with which the system can be expected to produce reliable results rather than explaining how an AI model arrives at a particular judgement (*166*). Clinicians require other types of information, even if they do not understand exactly how an algorithm functions, including the data on which it was trained, how and who built the AI model and the variables underlying the AI model.

Implications of the loss of human control in clinical care

Loss of human control by assigning decision-making to AI-guided technologies could affect various aspects of clinical care and the health-care system. They include the patient, the clinician–patient relationship (and whether it interrupts communication between them), the relation of the health-care system to technology providers and the choices that societies should make about standards of care.

Although providing individuals with more opportunities to share data and to obtain autonomous health advice could improve their agency and self-care, it could also generate anxiety and fatigue (*159*). As more personal data are collected by such technologies and used by clinicians, patients might increasingly be excluded from shared decision-making and left unable to exercise agency or autonomy in decisions about their health (*149*). Most patients have insufficient knowledge about how and why AI technologies make certain decisions, and the technologies themselves may not be sufficiently transparent, even if a patient is well informed. In some situations, individuals may feel unable to refuse treatment, partly also because the patient cannot speak with or challenge the recommendation of an AI-guided technology (e.g. a notion that the "computer knows best") or is not given enough information or a rationale for providing informed consent (*149*).

Hospitals and health-care providers are unlikely to inform patients that AI was used as a part of decision-making to guide, validate or overrule a provider. There is, however, no precedent for seeking the consent of patients to use technologies for diagnosis or treatment. Nevertheless, the use of AI in medicine and failure to disclose its use could challenge the core of informed consent and wider public trust in health care. This challenge depends on whether any of the reasons for obtaining informed consent – protection, autonomy, prevention of abusive conduct, trust, self-ownership, non-domination and personal integrity – is triggered by the use of AI in clinical care (*167*). See Box 4 for additional discussion on whether and how providers should disclose the use of AI for clinical care.

Box 4. Informed consent during clinical care

Consider use of an AI in a hospital to make recommendations on a drug and dosage for a patient. The AI recommends a particular drug and dosage for patient A. The physician does not, however, understand how the AI reached its recommendation. The AI has a highly sophisticated algorithm and is thus a black box for the physician. Should the physician follow the AI's recommendation? If patients were to find out that an AI or machine-learning system was used to recommend their care but no one had told them, how would they feel? Does the physician have a moral or even a legal duty to tell patient A that he or she has consulted an AI technology? If so, what essential information should the physician provide to patient A? Should disclosure of the use of AI be part of obtaining informed consent and should a lack of sufficient information incur liability? (*167*)

Transparency is crucial to promoting trust among all stakeholders, particularly patients. Physicians should be frank with patients from the onset and inform them of the use of AI rather than hiding the technology. They should try their best to explain to their patients the purpose of using AI, how it functions and whether it is explainable. They should describe what data are collected, how they are used and shared with third parties and the safeguards for protection of patients' privacy. Physicians should also be transparent about any weaknesses of the AI technology, such as any biases, data breaches or privacy concerns. Only with transparency can the deployment of AI for health care and health science, including hospital practice and clinical trials (*168*), become a long-term success. Trust is key to facilitating the adoption of AI in medicine.

Note: This case study was written by Marcelo Corrales Compagnucci (CeBIL Copenhagen), Sara Gerke (Harvard Law School) and Timo Minssen (CeBIL Copenhagen).

Physicians who are left out of decision-making between a patient and an AI health technology may also feel loss of control, as they can no longer engage in the back-and-forth that is currently integral to clinical care and shared decision-making between providers and patients (*160*). Some may consider loss of physician control over patients as promoting patient autonomy, but there is equally a risk of surrendering decision-making to an AI technology, which may be more likely if the technology is presented to the patient as providing better insight into their health status and prognosis than a physician (*160*).

Furthermore, if an AI technology reduces contact between a provider and a patient, it could reduce the opportunities for clinicians to offer health promotion interventions to the patient and undermine general supportive care, such as the benefits of human–human interaction when people are often at their most vulnerable (*159*). Some AI technologies do not sever the relationship between doctor and patient but help to improve contact and communication, for example, by providing an analysis of different treatment options, which the doctor can talk through with the patient and explain the risks.

Loss of control could be construed as surrendering not just to a technology but also to companies that exert power over the development, deployment and use of AI for health care. At present, technology companies are investing resources to accumulate data, computing power and human resources to develop new AI health technologies (*169–171*). This may be done by large companies in partnership with the public sector, as in the United Kingdom (*168*), but could be done by concentrating different areas of expertise or decision-making in different companies, with the rules and standards of care governed by the companies that manage the technologies rather than health care systems. In China, several large technology companies, including Ping An (*171*), Tencent (*174*), Baidu (*175*) and Alibaba (*176*), are rapidly expanding the provision of both online and offline health services and new points of access to health care, backed by accumulation of data and use of AI. Companies, unlike health systems or governments, may, however, ignore the needs of citizens and the obligations owed to citizens, as there is a distinction between citizens and customers. These concerns heighten the importance of regulation and careful consideration of the role of companies in direct provision of health-care services.

The ethics of using AI for resource allocation and prioritization

Use of computerized decision-support programs – AI or not – to inform or guide resource allocation and prioritization for clinical care has long raised ethical issues (*177*). They include managing conflicts between human and machine predictions, difficulty in assessing the quality and fitness for purpose of software, identifying appropriate users and the novel situation in which a decision for a patient is guided by a machine analysis of other patients' outcomes. In some situations, well-intentioned efforts to base decisions about allocations on an algorithm that relies only on a rules-based formula produce unintended outcomes. Such was the case in allocation of vaccines against COVID-19 at a medical institution in California, USA, on the basis of a rules-based formula in which very few of the available vaccine doses were allocated to those medical workers most at risk of contracting the virus, while prioritizing "higher-ranked" doctors at low-risk of COVID-19 (*178*).

Moreover, there is a familiar problem and risk that data in both traditional databases and machine-learning training sets might be biased. Such bias could lead to allocation of resources that discriminates against, for example, people of colour; decisions related to gender, ethnicity or socioeconomic status might similarly be biased. Such forms of bias and discrimination might not only be found in data but intentionally included in algorithms, such that formulas are written to discriminate against certain communities or individuals. At population level, this could encourage use of resources for people who will have the greatest net benefit, e.g. younger, healthier individuals, and divert resources and time from costly procedures intended for the elderly. Thus, if an AI technology is trained to "maximize global health", it may do so by allocating most

resources to healthy people in order to keep them healthy and not to a disadvantaged population. This dovetails with a wider "conceptual revolution" in medicine, whereas

> twentieth-century medicine aimed to heal the sick. Twenty-first-century medicine is increasingly aimed to upgrade the healthy.... Consequently, by 2070 the poor could very well enjoy much better healthcare than today, but the gap separating them from the rich will nevertheless be much greater (*179*).

As more data are amassed and AI technologies are increasingly integrated into decision-making, providers and administrators will probably rely on the advice given (while guarding against automation bias). Yet, such technologies, if designed for efficiency of resource use, could compromise human dignity and equitable access to treatment. They could mean that decisions about whether to provide certain costly treatments or operations are based on predicted life span and on estimates of quality-adjusted life years or new metrics based on data that are inherently biased. In some countries in which AI is not used, patients are already triaged to optimize patient flow, and such decisions often affect those who are disadvantaged or powerless, such as the elderly, people of colour and those with genetic defects or disabilities.

Ethical design (see section 7.1) could mitigate these risks and ensure that AI technologies are used to assist humans by appropriate resource allocation and prioritization. Furthermore, such technologies must be maintained as a means of aiding human decision-making and assuring that humans ultimately make the right critical life-and-death decisions by adequately addressing the risks of such uses of AI and providing those affected by such decisions with contestation rights.

Use of AI tools for triage or rationing is one of the most compelling reasons for ensuring adequate governance or oversight. Although intentional harm is not ethically controversial – it is wrong – the possibilities of unintended bias and flawed inference emphasize the need to protect and insulate people and processes from computational misadventure.

Use of AI for predictive analytics in health care

Health care has always included and depended in part on predictions and prognoses and the use of predictive analytics. AI is one of the more recent tools for this purpose, and many possible benefits of prediction-based health care rely on use of AI. AI could also be used to assess an individual's risk of disease, which could be used for prevention of diseases, such as heart disease and diabetes. AI could also assist health-care providers in predicting illness or major health events. For example, early studies with limited datasets indicated that AI could be used to diagnose Alzheimer disease years before symptoms appear (*180*).

Challenges to prediction in clinical care predate the emergence of AI and should not be attributed solely to AI techniques. Yet, various risks are associated with the use of AI to make predictions that affect patient care or influence the allocation of resources by a hospital or health-care system. Prediction technologies could be inaccurate because an AI technology bases its recommendations on an inference that optimizes markers of health rather than identifying an underlying patient need. An algorithm that predicts mortality from training data may have learnt that a patient who visits a chaplain is at increased risk of death (*181*).

While AI-based diagnosis is near term and its efficiency can be tested, thereby mitigating potential harm, efficacy and accuracy in long-term predictions may be more difficult or impossible to achieve. The risk of harm therefore increases dramatically, as predictions of limited reliability could affect an individual's health and well-being and result in unnecessary expenditure of scarce resources. For example, an AI-based mobile app developed by DeepMind to predict acute kidney failure produced two false-positive results for every correct result and therefore did not improve patient outcomes (*182*). Even if the system identified some patients who required treatment, this benefit was cancelled out by overdiagnosis. Such false-positive results can harm patients if they persuade doctors to take riskier courses of action, such as prescription of a more potent, addictive drug, in response to the prediction.

Prediction-based health care, even if it is effective for diagnosis or accurate prediction of disease, may present significant risks of bias and discrimination for individuals because of a predisposition to certain health conditions (*183*), which could manifest itself in the workplace, health insurance or access to health-care resources.
The use of predictions throughout health care also raises ethical concern about informed consent and individual autonomy if predictions are shared with people who did not consent to surveillance, detection or use of predictive models to draw inferences about their future health status or to provide them with a "predictive diagnosis" that they did not request in advance. Such non-consensual misuse could include, for example, screening to predict psychotic episodes by analysis of speech patterns (*184*) or use of AI to identify individuals with tuberculosis who do not know their status (as described above) or at high risk of HIV infection and thus candidates for pre-exposure prophylaxis (*185*). The Convention for the Protection of Human Rights and Dignity of the Human Being about the Application of Biology and Medicine (Oviedo Convention) (*68*) states that: "Everyone is entitled to know any information collected about his or her health. However, the wishes of individuals not to be so informed shall be observed."

Prediction-based technologies that are considered far more accurate or effective than older technologies could also challenge individual freedom of choice, even outside the doctor–patient relationship. Such use of AI, combined for example with "nudging",

could transform an application for promoting healthy behaviour into a technology that could exert powerful control over the choices people make in their daily lives (*105*), because nudging and the many ways in which it can be done can be far more effective than sporadic interactions between a health-care provider and a patient. If AI predicts that an individual is at high risk of a certain disease, will that individual still have the right to engage in behaviour that increases the likelihood of the disease? Such restrictions on autonomy could be imposed by a doctor but also by an employer or insurer or directly by an AI application on a wearable device.

Thus, while the introduction of prediction-based algorithms is often well-intentioned, the challenges and problems associated with their use can cause more harm than benefit, as was a predictive algorithm for assessing the likelihood of pregnancy in adolescents in vulnerable populations (Box 5).

Box 5. Challenges associated with a system for predicting adolescent pregnancy in Argentina

In 2017, the province of Salta, Argentina, signed an agreement with Microsoft to use AI to prevent adolescent pregnancy, a public health objective, and a tool to prevent school dropout. Microsoft used data for AI training collected by the local government from populations in vulnerable situations. The local authorities described the system (*186*) as

> intelligent algorithms that identify characteristics in people that can lead to some of these problems [adolescent pregnancy and school dropout] and warn the government so that they can work on prevention.

The data processed by Microsoft servers were distributed globally. It was claimed that, on the basis of the data collected, the algorithm would predict whether an adolescent would become pregnant with 86% accuracy (*187*). Once the partnership was publicized, however, it was challenged on technical grounds by local experts (*188*), for two reasons.

- Testing of the algorithms for predicting adolescent pregnancy had significant methodological shortcomings. The training data used to build the predictive algorithm and the data used to evaluate the algorithm's accuracy were almost identical, which gave rise to an erroneous conclusion about the predictive accuracy of the system.
- The type of data collected was inappropriate for ascertaining a future risk of pregnancy. The training data used were extracted from a survey of adolescents living in the province of Salta, which included personal information (e.g. age, ethnicity, country of origin), information about their environment (e.g. number of people in the household, whether they have hot water in the bathroom) and whether the person was pregnant at the time of the survey. These data were not appropriate for determining whether an individual would become pregnant in the future (e.g. within the ensuing 6 years), which would have required data collected 5 or 6 years before a pregnancy occurred. The collected data could be used at best only to determine whether an adolescent had been or was now pregnant.

The predictive algorithm was also inappropriate, as it provided predictions that were sensitive for adolescents without their (or their parents') consent, thereby undermining their privacy and autonomy. As the algorithm targeted individuals who were especially vulnerable, it was unlikely that they would have the opportunity to contest use of the interventions, and it could reinforce discriminatory attitudes and policies (*189*).

Despite the criticism and failings, the system continues to be used in at least two other countries (Brazil and Colombia) and in other provinces of Argentina (*187*). The flaws in the algorithm would have been identified more easily if there had been greater transparency about the data sets used to train and evaluate the algorithm, the technical specifications and the hypothesis that guided the model's design (*190*).

This case study was written by Maria Paz Canales (Derechos Digitales).

Use of AI for prediction in drug discovery and clinical development

It is expected that machine-learning systems will be used to predict which drugs will be safe and effective and are best suited for human use. Machine learning may also be used to design drug combinations to optimize the use of promising AI or conventionally designed drug candidates. Such predictive models could allow pharmaceutical companies to take "regulatory shortcuts" and conduct fewer clinical trials and with fewer patient data. A possible benefit of AI may therefore be to accelerate the development of medicines and vaccines, especially for new diseases with pandemic potential for which there are ineffective or no medical countermeasures.

Such approaches can, however, carry risks if AI is used incorrectly or too aggressively. Predictive models are based on algorithms that must be assessed for accuracy, which may be difficult because of lack of transparency or explainability about how the algorithms function. Furthermore, reducing the number of trials or patients studied can raise concern that patients may be exposed to risks that were not identified by the algorithm.

6.6 Bias and discrimination associated with artificial intelligence

Societal bias and discrimination are often replicated by AI technologies, including those used in the criminal justice system, banking, human resources and the provision of public services. The different forms of discrimination and bias that a person or a group of people suffer because of identities such as gender, race and sexual orientation must be considered. Racial bias (in the USA and other countries) is affecting the performance of AI technologies for health (Box 6).

Box 6. Discrimination and racial bias in AI technology

In a study published in Science in October 2019 (*191*), researchers found significant racial bias in an algorithm used widely in the US health-care system to guide health decisions. The algorithm is based on cost (rather than illness) as a proxy for needs; however, the US health-care system spent less money on Black than on white patients with the same level of need. Thus, the algorithm incorrectly assumed that white patients were sicker than equally sick Black patients. The researchers estimated that the racial bias reduced the number of Black patients receiving extra care by more than half.

This case highlights the importance of awareness of biases in AI and mitigating them from the onset to prevent discrimination (based on e.g. race, gender, age or disability). Biases may be present not only in the algorithm but also, for example, in the data used to train the algorithm. Many other types of bias, such as contextual bias (*192*, *193*), should be considered. Stakeholders, particularly AI programmers, should apply "ethics by design" and mitigate biases at the outset in developing a new AI technology for health (*194*).

Note: This case study was written by Marcelo Corrales Compagnucci (CeBIL Copenhagen), Sara Gerke (Harvard Law School) and Timo Minssen (CeBIL Copenhagen).

Bias in data

The data sets used to train AI models are biased, as many exclude girls and women, ethnic minorities, elderly people, rural communities and disadvantaged groups. In general, AI is biased towards the majority data set (the populations for which there are most data), so that, in unequal societies, AI may be biased towards the majority and place a minority population at a disadvantage. Such systematic biases, when enshrined in AI, can become normative biases and can exacerbate and fix (in the algorithm) existing disparities in health care (*195*). Such bias is generally present in any inferential model based on pattern recognition. Thus, the human decisions that

> comprise the data and shape the design of the algorithm [are] now hidden by the promise of neutrality and [have] the power to unjustly discriminate at a much larger scale than biased individuals (*196*).

Existing bias and established discrimination in health-care provision and the structures and practices of health care are captured in the data with which machine-learning models are trained and manifest in the recommendations made by AI-guided technologies. The consequence is that the recommendations will be irrelevant or inaccurate for the populations excluded from the data (Box 7), which is also the consequence of introducing an AI technology that is trained for use in one context into a different context.

Box 7. AI technologies for detecting skin cancer exclude people of colour.

Machine learning has outperformed dermatologists in detecting potentially cancerous skin lesions. As rates of skin cancer increase in many countries, AI technology would improve the ability of dermatologists to diagnose skin cancer. The data used to train one highly accurate machine-learning model are, however, for "fair-skinned" populations in Australia, Europe and the USA. Thus, while the technology assists in diagnosis, prevention and treatment of skin cancer in white and light-skinned individuals, the algorithm was neither appropriate nor relevant for people of colour, as it was not trained on images of these populations.

The inadequacy of the data on people of colour is due to several structural factors, including lack of medical professionals and of adequate information in communities of colour and economic barriers that prevent marginalized communities from seeking health care or participating in research that would allow such individuals to contribute data.

Another reason that such machine-learning models are not relevant for people of colour is that developers seek to bring new technologies to the market as quickly as possible. Even if their haste is guided by a desire to reduce avoidable morbidity and mortality, it can replicate existing racial and ethnic disparities, while a more deliberate, inclusive approach to design and development would identify and avoid biased outcomes.

Source: reference 197

Such biases in data could also affect, for example, the use of AI for drug development. If an AI technology is based on a racially homogenous dataset, biomarkers that an AI technology identifies and that are responsive to a therapy may be appropriate only for the race or gender of the dataset and not for a more diverse population. In such cases, a drug that is approved may not be effective for the excluded population or may even be harmful to their health and well-being.

Data biases are also due to other factors. One is the digital divide. (See section 6.2.) Thus, women in LMIC are much less likely than men to have access to a mobile phone or mobile Internet; 327 million fewer women than men have access to mobile Internet (*198*). Thus, women not only contribute fewer data to data sets used to train AI but are less likely to benefit from services. Another cause is unbalanced collection of data, even where the digital divide is not a factor. For example, genetic data tend to be collected disproportionately from people of European descent (*199*, *200*). Furthermore, experimental and clinical studies tend to involve male experimental models or male subjects, resulting in neglect of sex-specific biological differences, although this gap may be closing slightly (*201*).

Biases can also emerge when certain individuals or communities choose not to provide data. Data on certain population subsets may be difficult to collect if collection requires expensive devices such as wearable monitors. As noted above, improving data collection from such communities or individuals, while it may improve the performance of AI, carries a risk of data colonialism. (See section 6.3.)

Biases related to who develops AI and the origin of the data on which AI is trained

Biases often depend on who funds and who designs an AI technology. AI-based technologies have tended to be developed by one demographic group and gender, increasing the likelihood of certain biases in the design. Thus, the first releases of the Apple Health Kit, which enabled specialized tracking of some health risks, did not include a menstrual cycle tracker, perhaps because there were no women on the development team (*202*).

Bias can also arise from insufficient diversity of the people who label data or validate an algorithm. To reduce bias, people with diverse ethnic and social backgrounds should be included, and a diverse team is necessary to recognize flaws in the design or functionality of the AI in validating algorithms to ensure lack of bias.
Bias may also be due to the origin of the data with which AI is designed and trained. It may not be possible to collect representative data if an AI technology is initially trained with data from local populations that have a different health profile from the populations in which the AI technology is used. Thus, an AI technology that is trained in one country and then used in a country with different characteristics may

discriminate against, be ineffective or provide an incorrect diagnosis or prediction for a population of a different race, ethnicity or body type. AI is often trained with local data to which a company or research organization has access but sold globally with no consideration of the inadequacy of the training data.

Bias in deployment

Bias can also be introduced during implementation of systems in real-world settings. If the diversity of the populations that may require use of an AI system, due to variations in age, disability, co-morbidities or poverty, has not been considered, an AI technology will discriminate against or work improperly for these populations. Such bias may manifest itself at the workplace, in health insurance or in access to health-care resources, benefits and other opportunities. As AI is designed predominantly in HIC, there may be significant misunderstanding of how it should be deployed in LMIC, including the discriminatory impact (or worse) or that it cannot be used for certain populations.

6.7 Risks of artificial intelligence technologies to safety and cybersecurity

This section discusses several risks for safety and cybersecurity associated with use of AI technologies for health, which may be generalized to the use of many computing technologies for health care – past and present.

Safety of AI technologies

Patient safety could be at risk from use of AI that may not be foreseen during regulatory review of the technology for approval. Errors in AI systems, including incorrect recommendations (e.g. which drug to use, which of two sick patients to treat) and recommendations based on false-negative or false-positive results, can cause injury to a patient (*159*) or a group of people with the same health condition. Model resilience, or how an AI technology performs over time, is a related risk. Health-care providers also commit errors of judgement and other human errors, but the risk with AI is that such an error, if fixed in an algorithm, could cause irreparable harm to thousands of people in a short time if the technology is used widely (*159*). Furthermore, the psychological burden and stress of such errors is borne by the providers who operate such technologies.

An AI application, like any information technology system, could also provide the wrong guidance if it has code errors due to human programming mistakes. For example, the United Kingdom NHS COVID-19 application, which was designed to notify individuals to self-isolate if exposed, was programmed incorrectly (*203*). Thus, a user of the application had to be next to a highly infectious patient five times longer than that considered risky by the NHS before being instructed to self-isolate. Although up to 19 million people downloaded the application, a "shockingly low" number of people were told to isolate, thereby exposing themselves and others to risks of COVID-19 infection (*203*).

It is also possible that a developer (or an entity that funds or directs the design of AI technology) designs an AI technology unethically, to optimize an outcome that would generate profits for the provider or conceal certain practices. The design might in fact be more accurate than another modelling technique but generate unmerited sales revenue. Malicious design has affected other sectors, such as the automobile sector, in which algorithms used to measure emissions were programmed to conceal the true emissions profile of a major car manufacturer (*204*).

Use of computers carries an inherent risk of flaws in safety due to insufficient attention to minimizing risk in the design of machines and also to flaws in the computer code and associated bugs and glitches. Injuries and deaths due to such flaws and breakdowns are underreported, and there are no official figures and few large-scale studies. In one study in the United Kingdom, for instance, it was estimated that up to 2000 deaths a year may be due to computer errors and flaws and that it is an "unnoticed killer" (*205*).

Cybersecurity

As health-care systems become increasingly dependent on AI, these technologies may be expected to be targeted for malicious attacks and hacking in order to shut down certain systems, to manipulate the data used for training the algorithm, thereby changing its performance and recommendations, or to "kidnap" data for ransom (*181*). AI developers might be targeted in "spear-fishing" attacks and by hacking, which could allow an attacker to modify an algorithm without the knowledge of the developer.

An algorithm, especially one that runs independently of human oversight, could be hacked to generate revenue for certain recipients, and large sums are at stake: total spending on health care globally was US$ 7.8 trillion in 2017, or about 10% of global gross domestic product (*206*). The United Kingdom Information Commission Office noted that cyberattacks on the health sector are the most frequent (*207*). Breaches of health data, which are some of the most sensitive data about individuals, could harm privacy and dignity and the broader exercise of human rights. A study in 2013 showed that four anonymized data points are sufficient for unique identification of an individual with 95% accuracy (*208*). Measures to avoid such breaches, which can be broadly categorized as infrastructural or algorithmic, are improving, although no defence is 100% effective and new defences can be broken as quickly as they are proposed (*181*).

6.8 Impacts of artificial intelligence on labour and employment in health and medicine

The impact of AI on the health workforce is viewed with equal optimism and pessimism. It is perhaps less contested that nearly all jobs in health care will require a minimum level of digital and technological proficiency. The Topol Review: Preparing the health workforce to deliver the digital future (*24*), concluded that, within two decades,

90% of all jobs in the United Kingdom's NHS will require digital skills, including navigating the "data-rich" health-care environment, and also digital and genomics literacy. The requirement for digital literacy will not be limited to clinical care (although this section concentrates on clinical staff) but extends to health-care workers in public health, surveillance, the environment, prevention, protection, education, awareness, diet, nutrition and all the other social determinants of health that can be supported by AI. All health workers in these areas will have to be trained and retrained in use of AI to support and facilitate their tasks.

Optimistic views include that in which AI will automate and thus reduce the burden of routine tasks on clinicians and allow them to focus on more challenging work and to engage with patients. It could also empower doctors to work in more areas and provide support in areas in which technology can be used for clinical decision-making. It is expected that digitization of health care and the introduction of AI technologies will create numerous new jobs in health care, such as software development, health-care systems analysis and training in the use of AI for health care and medicine. The last may include three types of jobs: trainers, or people that can evaluate and stress-test AI technologies; explainers, or those who can explain how and why an algorithm can be trusted; and "sustainers", or those who monitor behaviour and identify unintended consequences of AI systems (*181*).

AI could also extend one of the scarcest resources in health-care systems – the time that doctors and nurses have to attend to patients. If doctors and nurses can hand over repetitive or administrative tasks to AI-supported technologies and therefore spend less time on "routine care cases", they would have more time to attend to more urgent, complex or rare cases and to improve the overall quality of care offered to patients (*24*). In some cases, however, as AI is being integrated into health-care systems as secondary medical support, during what could best be described as a transition period, AI may increase the tasks and add to the workload of doctors and nurses.

Telemedicine has been used to extend health-care provision to people in remote areas and to refugees and other underserved populations that otherwise lack appropriate medical advice (*205*). Yet, AI and its use in telemedicine could create inequitable access to health-care services (in particular to health-care personnel), for instance when people in rural areas or low-income countries have to make do with greater access to AI-based services and telemedicine (*181*) while individuals in HIC and urban areas continue to benefit from in-person care.

Furthermore, health-care workers who already have to absorb large amounts of information to meet standards of care may regularly require new competence in the use of AI-supported technologies in everyday practice, and competence may have to evolve rapidly as the uptake of AI accelerates. Such continuing education may be

neither available nor accessible to all health-care workers, although efforts are under way to improve digital literacy and training that includes use of AI and other health information technologies. (See section 7.2.)

Even as health-care workers have to obtain new competence, the use of AI to augment and possibly replace the daily tasks of health-care workers and physicians could also remove the need for maintaining certain skills, such as the ability to read an X-ray. At some point, physicians may be unable to conduct such a task without the assistance of a computer, and AI systems will have to be "trained" to use the repository of medical knowledge that was the domain of human providers (*159*). Such dependence on AI systems could erode independent human judgement and, in the worst-case scenario, could leave providers and patients incapable of acting if an AI system fails or is compromised (*159*). There should therefore be robust plans to provide back-up if technology systems fail or are breached.

Another concern is that AI will automate many of the jobs and tasks of health-care personnel, resulting in significant loss of jobs in nearly every part of the health workforce, including certain types of doctors. AI has already replaced many jobs in other industries, reduced the total number of people required for certain roles or created the expectation that many jobs will be lost (e.g. up to 35% of all jobs in the United Kingdom) (*210*).

In many countries, however, health care is not an industry but a core government function, so that administrators will not replace health-care workers with technology. Many countries, with high, middle or low income, are in fact facing shortages of health-care workers. WHO has estimated that, by 2030, there will be a shortage of 18 million health workers, mostly in low- and low- to middle-income countries (*211*). AI may provide a means to bridge the gap between the workforce ideally available to provide appropriate health care and what exists.

Other scenarios have been envisaged with the arrival of AI. One predicts that a decision to use AI will cause short-term instability, with many job losses in certain areas even as overall employment increases with the creation of new jobs, resulting in unemployment for those who may not be able to retrain for the new roles. In another scenario, job losses will not materialize, either because clinicians or health-care workers will fulfil other roles or because these technologies will be fully integrated only over a long time, during which other roles for health-care workers and clinicians will emerge, such as labelling data or designing and evaluating AI technologies (*210*).

Even if AI does not displace clinicians, it could make doctors' jobs less secure and stable. One trend has been the "Uberization" of health care, in which AI facilitates the creation of health-care platforms on which contractors, including drivers, temporary workers,

nurses, physician assistants and even doctors, work on demand (*103*, *211*). During the past decade, health care and education have seen the fastest growth of "gig workers", who work on a temporary basis with no stability of employment (*103*). While this provides more flexible services, it could also sever relationships between patients and health-care givers and create insecurity for certain types of health workers. Such a trend may not occur in countries with either greater labour protection for its health workforce, such that labour shortages provide health-care workers with negotiating power, or in which AI is not used to reorganize health care but to reduce the workload.

With increasing use of AI, the nature of medical practice and health-care provision will fundamentally change. As noted above, it could provide health-care workers with more time to care for patients or it could, if patients interact more frequently and directly with AI, result in doctors spending less time in direct contact with patients and more time in administering technology, analysing data and learning how to use new technologies. If introduction of AI is not effectively managed, physicians could become dissatisfied and even leave medical practice (*213*).

6.9 Challenges in commercialization of artificial intelligence for health care

There are various ethical challenges to the practices of the largest technology firms in the field of AI for health, although some of the concerns also apply to mid-size firms and start-ups. The use of AI for health has been pushed by companies – from small start-up firms to large technology companies – mainly by significant advocacy and investment. Those who support a growing role for these companies expect that they will be able to marshal their capital, in-house expertise, computing resources and data to identify and build novel applications to support providers and health systems. During the COVID-19 pandemic, many companies have sought to provide services and products for the response, many of which are linked to forms of public health surveillance (*214*). This has raised a number of ethical and legal concerns, which are discussed throughout this report.

Some services already widely used in health are for "back-office" functions and for managing health-care systems. Some of the companies involved in development of technology, such as the pharmaceutical and medical device industries, are integrating AI into their processes and products, and insurance firms are using AI for assessing risk or even automating the provision of insurance, which might raise ethical concerns with respect to algorithmic decision-making.

A prominent use of AI for health care is to support diagnosis, treatment, monitoring and adherence to treatment. Such applications could have benefits for health-care systems; however, many concerns have emerged during the past as more technology firms, and especially the largest firms, have entered the health-care field.

A general problem is lack of transparency. While many firms know much about their users, their users, civil society and regulators know little about the activities of the firms, including how they (and governments) operate in PPPs, which have a significant impact on the public interest (*215*). (See section 9.3.) Their practices remain hidden partly because of commercial secrecy agreements or the lack of general obligations for transparent practices, including the role these firms play in health care and the data that are collected and used to train and validate an AI algorithm. Without transparency (and accountability), these firms have little incentive to act in a way that does not cross certain ethical boundaries or to disclose deeper problems in their technology, data or models (*215*). Many companies prefer to keep their algorithmic models proprietary and secret, as full transparency could lead to criticism of both the technology and the company (*216*).

A second broad concern is that the overall business model of the largest technology firms includes both aggressive collection and use of data to make their technologies effective and use of surplus data for commercial practices, considered by Professor Shoshana Zuboff as "surveillance capitalism" (*125*). Thus, during the past decade, there have been several examples of large technology firms using large datasets of sensitive health information in developing AI technologies for health care (*129*, *217*). While such health data may have been acquired and used to develop useful AI technologies for health, the data were not acquired with the explicit consent of those who provided them, the benefits of the data for these firms may be far in excess of what was required to deliver the product, and the firms may not provide equal benefits to the population that generated the data in the first place.

Such acquisition of sensitive health information can give rise to legal concern. First, even if the data are anonymized by the firm that acquires them, the company would be able to combine data and de-anonymize relevant data sets from the amount of information it already has from other sources (*147*). Secondly, several large technology firms have been accused and even fined for mishandling data (*218*), and this concern may be heightened for firms that acquire often-sensitive health data. Thirdly, as firms continue to accumulate large amounts of data, this can introduce anti-trust concerns (although it may not lead to regulatory enforcement (*219*)), related to the growing market power of such companies, including barriers to smaller companies that may wish to enter an AI market (*220*).

An additional concern is the growing power that some companies may exert over the development, deployment and use of AI for health (including drug development) and the extent to which corporations exert power and influence over individuals and governments and over both AI technology and the health-care market. Data, computing power, human resources and technology can be concentrated within a few companies, and technology can be owned either legally (IP protection) or because the size of a company's platform results in a monopoly. Monopoly power can concentrate

decision-making in the hands of a few individuals and companies, which can act as gatekeepers of certain products and services (*221*) and reduce competition, which could eventually translate into higher prices for goods and services, less consumer protection or less innovation.

While the growing role of large companies in the USA, such as Google, Facebook and Amazon, in the development and provision of AI for health care has been under scrutiny, large technology companies in China and other Asian countries are playing a similar role in health through such services and technologies. They include Ping An, Tencent, Baidu and Alibaba, which are both building their own technology platforms and collaborating with user platforms such as WeChat to reach millions of people in China (*176*). Tencent, for example, is investing in at least three main areas of health: AI-based technologies to assist in diagnosis and treatment, a "smart hospital" to provide a web of online services and data connectivity through a smart health card (which itself raises concern about data privacy and use; see above) and a "medipedia" to provide health information to users online (*222*). Alibaba is working with hospitals to predict patient demand in order to allocate health-care personnel and developing AI-assisted diagnostic tools for radiology (*176*).

Such power and control of the market by large firms may be part of a 'first-mover' advantage that several large firms may eventually earn through their entry into AI for health. Even if the data used by a firm (for example, data from a public health system) could be used by others, other firms might be discouraged or unable to replicate use of such data for a similar purpose, especially if another company has already done so (*215*). Such power also means that the rules set by certain companies can force even the largest and wealthiest governments to change course. For example, during the COVID-19 pandemic, Google and Apple introduced a technical standard for where and how data should be stored in proximity-tracking applications that differed from the approach preferred by the governments of several HIC, which resulted in at least one government changing the technical design of its proximity-tracking application to comply with the technical standards of these two companies. Although the approach of these companies may have been consistent with privacy considerations, the wider concern is that these firms, by controlling the infrastructure with which such applications operate, can force governments to adopt a technical standard that is inconsistent with its own public policy and public health objectives (*223*).

When most data, health analytics and algorithms are managed by large technology companies, it will be increasingly likely that those companies will govern decisions that should be taken by individuals, societies and governments, because of their control and power over the resources and information that underpins the digital economy (*124*). This power imbalance also affects people who should be treated equitably by their governments or at least, if treated unfairly, can hold their governments

accountable if inequity arises. Without a strong government role, companies might ignore the needs of individuals, particularly those at the margins of society and the global economy (*179*).

Stringent oversight by governments and good governance are essential in this sector. (See section 9.3 on private sector governance.) Oversight mechanisms could be integrated into PPPs. If such partnerships are not carefully designed, they can lead to misappropriation of resources (usually patient data) or conflicts of interest in decision-making in such partnerships or could forestall or limit the use of regulation to protect the public interest when necessary (*215*, *216*).

6.10 Artificial intelligence and climate change

Use of deep learning models in AI has been scrutinized for its impact on climate change. Researchers at the University of Massachusetts Amherst, USA, found that the emissions associated with training a single "big language" model were equal to approximately 300 000 kg of carbon dioxide or 125 round-trip flights between New York City and Beijing (*224*). A single training session for another deep-learning model, GTP-3, requires energy equivalent to the annual consumption of 126 Danish homes and creates a carbon footprint equivalent to travelling 700 000 km by car (*225*). All the infrastructure required to support use of AI has an additional carbon cost (*225*).

WHO considers climate change to be an urgent, global health challenge that requires prioritized action now and in the decades to come. Between 2030 and 2050, climate change is expected to cause approximately 250 000 additional deaths per year from malnutrition, malaria, diarrhoea and heat stress alone. The cost of direct damage to health by 2030 is estimated to be US$ 2–4 billion per year. Areas with weak health infrastructure – most in developing countries – will be the least able to cope without assistance to prepare and respond (*226*).

Reducing emissions of greenhouse gases through better transport, food and choices of energy, particularly reducing air pollution, results in better health (*226*). Extending the use of AI for health and in other sectors of the global economy could, however, contribute directly to dangerous climate change and poor health outcomes, especially of marginalized populations. Thus, the growing success and benefits for health outcomes of AI, which will predominate in HIC, would be directly linked to increased carbon emissions and negative consequences in low-income countries.
AI technologies, for health and other uses, should therefore be designed and evaluated to minimize carbon emissions, such as by using smaller, more carefully curated data sets, which could also potentially improve the accuracy of AI models (*227*). Otherwise, the growing use of AI might have to be balanced against its impact on carbon emissions.

7. BUILDING AN ETHICAL APPROACH TO USE OF ARTIFICIAL INTELLIGENCE FOR HEALTH

This section addresses how measures other than law and policy can ensure that AI improves human health and well-being.

7.1 Ethical, transparent design of technologies

Although technology designers and developers play critical roles in designing AI tools for use in health, there are no procedures for credentialing or licensing such as those required for health-care workers. In the absence of formal qualifications for ethics in the AI field, it is not enough merely to call for personal adherence to values such as reproducibility, transparency, fairness and human dignity.

New approaches to software engineering in the past decade move beyond an appeal to abstract moral values, and improvements in design methods are not merely upgraded programming techniques. Methods for designing AI technologies that include moral values in health and other sectors have been proposed to support effective, systematic, transparent integration of ethical values. Such values in design have also been codified legally; for example, the GDPR includes specific obligations to include privacy by design and by default.

One approach to integrating ethics and human rights standards is "Design for values", a paradigm for basing design on the values of human dignity, freedom, equality and solidarity (Box 8) and for construing them as non-functional requirements (*228*). This requires not a solutions-oriented approach but instead a process-oriented approach that satisfies stakeholder needs in conformity with the moral and social values embodied by human rights.

Box 8. Design for values (*229*)

"Design for values" is explicit transposition of moral and social values into context-dependent design requirements. It is an umbrella term for several pioneering methods, such as value-sensitive design, values in design and participatory design. Design for values presents a roadmap for stakeholders to translate human rights into context-dependent design requirements through a structured, inclusive, transparent process, such that abstract values are translated into design requirements and norms (properties that a technology should have to ensure certain values), and the norms then become a socio-technical design requirement. The process of identifying design requirements permits all stakeholders, including individuals affected by the technology, users, engineers, field experts and legal practitioners, to debate design choices and identify the advantages and shortcomings of each choice.

Thus, a value such as privacy can be interpreted through certain norms, such as informed consent, right to erasure and confidentiality. These norms can then be converted by discussion and consultation into design requirements, such as positive opt-in (a means of ensuring informed consent) or homomorphic encryption techniques to assure confidentiality. Other techniques for safeguarding privacy, such as k-anonymity, differential privacy and coarse graining through clustering, could also be selected through consultation.

Ethical design can also be applied to the socio-technical systems in which algorithms are developed, which comprise the ensemble of software, data, methods, procedure, personnel, protocols, laws, norms, incentive structures and institutional frameworks. All are brought together to ensure that products and services provide ethical outcomes for society and its health-care systems.

More generally, ethical and transparent design of AI technologies should be ensured by prioritizing inclusivity in processes and methods (*230*, *231*). Consideration of inclusivity when designing and developing an AI technology can overcome barriers to equitable use of the technology in health associated with geography, gender, age, culture, religion or language.

Three approaches for promoting inclusivity are the following.

- *Citizen science*: Citizen science is defined by the Alan Turing Institute as the direct contribution of non-professional scientists to scientific research, for instance, by contributing data or performing tasks (*232*). Citizen science not only helps the public to understand a particular study or technology that may affect them personally but also ensures that the public is involved in research, discussions and tool-building. This ensures respectful co-creation of AI technologies that reduces the distance between the researcher or programmer and the individuals who the technology is intended to serve.

- *Open-source software:* Transparency and participation can be increased by the use of open-source software for the underlying design of an AI technology or making the source code of the software publicly available. Open-source software is open to both contributions and feedback, which allows users to understand how the system works, to identify potential issues and to extend and adapt the software. Open-source software design must be accessible and welcoming, and the content should allow greater engagement and transparency.
- *Increased diversity:* Too often, efforts to increase the diversity of AI technologies involve increasing the diversity of the data on which they are based. Although this is necessary, it is not sufficient and might even amplify any biases inherent in the design. Minimizing and identifying potential biases requires greater involvement of people who are familiar with the nature of potential biases, contexts and regulations throughout software development, from its design to consultation with stakeholders, labelling of data, testing and deployment.

Toolkits can be useful for providing concrete guidance to technology designers who wish to integrate ethical considerations into their work. Software developer kits can provide guidelines that include a code of ethics, with specific guidelines for health. Such kits could indicate, for example, how to manage data, including collection, de-identification and aggregation, and how to safeguard the destination of data.

Kits have also been developed to facilitate certain ethical (and increasingly legal) requirements, such as the Sage Bionetworks toolkit for the elements of informed consent (*233*). The toolkit provides use cases to explain its approach to informed consent, including eConsent, examples of how it should be put into practice, a checklist to ensure that programmers have considered all the necessary questions and additional resources.

With the proliferation of use of AI for health, the emergence of more not-for-profit AI developers would be beneficial. Such developers, who are not constrained by internal or external revenue targets, can adhere to ethical principles and values more readily than private developers. Not-for-profit developers may include treatment providers, hospital systems and charities. They could emulate the many partnerships for not-for-profit product development that have been formed during the past two decades in the development of new medicines, diagnostics and vaccines. The partnerships are often with the public and private sectors and focus on neglected populations while ensuring affordability and access to all. A not-for-profit developer could address all areas of health but particularly areas of neglect, while ensuring that their technologies adhere to ethical values such as privacy, transparency and avoidance of bias.

Putting prediction to good use

Use of AI for prognosis will allow assessment of the relative risk of disease and predict illness. There are, however, several risks and challenges with the use of predictive analytics, including concern about the accuracy of the predictions and that prediction of a negative outcome could affect an individual's autonomy and well-being.

In public health, predictive analytics can forecast major health events, including outbreaks, before they occur. For example, before the COVID-19 pandemic, WHO was developing EPI-BRAIN, a global platform that will allow experts in data and public health to analyse large datasets for use in emergency preparedness and response (*234*). It would allow forecasting and early detection of threats of infection and their impact on the basis of scenarios, simulation exercises and insights to improve coordinated decision-making and response.

Ethical, transparent design allows governments and international health agencies, such as WHO, to encourage the development of AI technologies for predictive analytics to assist and augment decision-making by providers and policy-makers. Such technologies must adhere to ethical standards and human rights obligations, should be open to improvement and should be available for adaptation and use by governments and providers on a non-exclusive basis.

Recommendations

1. Potential end-users and all direct and indirect stakeholders should be engaged from the early stages of AI development in structured, inclusive, transparent design and given opportunities to raise ethical issues, voice concerns and provide input for the AI application under consideration. Relevant ethical considerations should inform the design and translation of moral values into specific context-dependent design requirements.

2. Designers and other stakeholders should ensure that AI systems are designed to perform well-defined tasks with the accuracy and reliability necessary to improve the capacity of health systems and advance patient interests. Designers and other stakeholders should also be able to predict and understand potential secondary outcomes.

3. Designers should ensure that stakeholders have sufficient understanding of the task that an AI system is designed to perform, the conditions necessary to ensure that it can perform that task safely and effectively and conditions that might degrade system performance.

4. The procedures that designers use to "design for values" should be informed and updated by the consensus principles stated in this report, best practices (e.g. privacy preserving technologies and techniques), standards of ethics by design and evolving professional norms (transparency of access to codes, processes that allow verification and inclusion).

5. Continuing education and training programmes should be available to designers and developers to ensure that they integrate evolving ethical considerations into design processes and choices. The establishment of formal accreditation procedures could ensure that designers and developers abide by ethical principles similar to those required of health-care workers.

7.2 Engagement and role of the public and demonstration of trustworthiness to providers and patients

Effective use of AI for health will require building the trust of the public, providers and patients. Social license requires hard-fought efforts that can be surrendered quickly if AI technologies are introduced without due care for the perspectives of those affected by its use. Public engagement and dialogue are means to ensure that use of AI for health care meets certain core societal expectations and greater trust and acceptance. Public dialogue also allows ascertainment of society's views, as far as possible, on the ethical dimensions of AI, its design and uses.

A critical issue of public concern, discussed throughout this publication, is the collection and use of patient data for AI and other applications. In the United Kingdom, these concerns have been addressed in public debate and dialogue. Health Data Research, which collects health data and makes it available to public and private entities for health-related applications of AI,[7] has used public engagement, including with the Wellcome Trust's initiative, Understanding Patient Data (*236*). Workshops held as part of the initiative provided a forum for participants to discuss their expectations and concerns about use of patient data in AI and other applications. Before these workshops, 18% of participants considered it acceptable to share anonymized patient data with commercial organizations for reasons other than direct care; after the workshops, the proportion had increased to 45% (*237*). Individuals who expressed positive views considered that contributing data was a value exchange, with a societal benefit, and wanted the NHS to benefit from their data. They also considered it acceptable for commercial companies to have access to their data, provided that the benefit returned to the public and that the NHS administered the data for the public benefit.

The United Kingdom Academy of Medical Sciences found at its meetings and workshops (*238*) that:

> ongoing engagement with patients, the public and healthcare professionals, including via co-creation, will be critical to ensuring new AI technologies respond to clinical unmet need, are fit for purpose, and are successfully deployed, adopted and used.

The Academy conducted a public dialogue on the "data-driven future" to understand awareness, expectations, aspirations and concerns about future technologies that

[7] Presentation by Dr Andrew Morris, Health Data Research United Kingdom, 3 October 2019 to the WHO working group on ethics and governance of AI for health.

would require patient data to be accessed, analysed or linked for clinical diagnosis and management (*239*). The respondents considered that any new use of data must have a proven social benefit and that an appropriate organization (such as the government or the NHS) should oversee the data and administer it for the public benefit.

Steps must be taken to build the trust of providers and patients who will increasingly rely on AI for routine clinical decision-making. The willingness of patients to rely on AI may sometimes be much lower than expected. For example, in a study conducted by HSBC Bank (*240*), only 8% of the respondents surveyed said that they would trust a machine offering mortgage advice, while 41% said they would trust a mortgage broker. Lack of wider trust could create significant divisions in a health-care system, in which, for example, older patients might be unwilling to adapt to and use new AI technologies, while younger patients might be more amenable (*155*).

With such a low level of trust, scandals that emerge from use of AI for health care and undermine patients' economic, personal or physical security could be fatal. After the Cambridge Analytica scandal in 2019, an estimated 15% of Facebook users surveyed indicated they would reduce their use of the social networking site. Trust could be eroded even more quickly and severely in the domain of health care if similar scandals or abuses of trust emerged into public discourse, destroying public trust overnight (*158*).

One means of mitigating and managing risk would be to allow health-care providers and developers to test a new AI product or service in a "live environment" in a testing facility, with safeguards and oversight to protect the health system from any risks or unintended consequences. Testing facilities could allow assessment, certification and validation of AI. In limited circumstances, testing facilities could build a "regulatory sandbox" (*241*), which might, however, be appropriate only in countries in which new health-care products and services and their specifications are subject to formal regulation and to data protection regulations (*242*). Examples of the use of regulatory sandboxes are the United Kingdom's Care Quality Commission and by the Singapore Government to test new digital health models (*242*).

A second approach to building trust and facilitating a "graceful transition" of health care is to redesign training programmes for the health workforce (Box 9) and to improve general education (*243*). Improvements in general education would include primary education in science, technology and mathematics.

Box 9. Supporting health workers in the use AI technologies, including through education and training

Medical professionals and health-care workers should receive sufficient technical, managerial and administrative support, capacity-building, regulatory protection (when appropriate) and training in the many uses of AI technologies and their advantages and in navigating the ethical challenges of AI (*244*). With regard to education and training, AI curricula should be seamlessly integrated into existing programmes (*244*). Curricula should be updated regularly, as AI is evolving continuously. Some members of the health-care profession will require training in basic use of computers before they adapt to use of AI. All health-care professionals will require a certain level of digital literacy, defined in the Topol review as "those digital capabilities that fit someone for living, learning, working, participating and thriving in a digital society" (*24*).

Physicians and nurses will also require a wider range of competence to apply AI in clinical practice, including better understanding of mathematical concepts, the fundamentals of AI, data science, health data provenance, curation, integration and governance (*24*), and also of the ethical and legal issues associated with the use of AI for health. Such measures (including training) will be necessary to combine and analyse data from many sources adequately, supervise AI tools and detect inaccurate performance of AI (*244*). Good support and training will ensure that health-care workers and physicians, for example, can avoid common pitfalls such as automation bias when using AI technologies. Eventually, the knowledge, skills and capabilities required of health workers may be defined by professional and statutory regulatory bodies in collaboration with practitioners and educators (*24*).

Significant changes may be made to medical education. Rather than rote memorization, which has been the hallmark of medical training, medical students might instead build and refine their competence for communication and negotiation, emotional intelligence, the ability to resolve ethical dilemmas and proficient use of computers. Medical training programmes will therefore require new educators who can teach these concepts and skills (*24*).

A third approach, the use of human warranty, is discussed earlier in this report (section 5), whereby developers of AI technologies work directly with providers and patients in patient and clinical evaluation at critical points in the development and deployment of the technologies. Human warranty can ensure meaningful public consultation and debate (*101*).

Recommendations

1. The public should be engaged in the development of AI for health in order to understand forms of data sharing and use, to comment on the forms of AI that are socially and culturally acceptable and to fully express their concerns and expectations. Further, the general public's literacy in AI technology should be improved to enable them to determine which AI technologies are acceptable.

2. Training and continuing education programmes should be available to assist health-care professionals in understanding and adapting to use of AI, learning about its benefits and risks and understanding the ethical issues raised in their use.

7.3 Impact assessment

An impact assessment is used to predict the consequences of a current or proposed action, policy, law, regulation or, as in the case of use of AI for health, a new technology or service. Impact assessments can provide both technical information on possible consequences and risks (both positive and negative) and improve decision-making, transparency and participation of the public in decision-making and introduce a framework for appropriate follow-up and measurement. Such assessment might be especially important for the use of AI, as an AI technology can change over time (*245*). Impact assessments can also be used to determine whether a technology will respect or undermine ethical principles and human rights obligations, including privacy and non-discrimination. Several types of impact assessment for the use of AI for health have been proposed or used, which could be considered by governments, companies and providers.

Businesses that design and introduce AI technologies for health have a particular obligation to conduct impact assessments, including on human rights. The UN Guiding Principles on Business and Human Rights of the United Nations Office of the High Commissioner for Human Rights establish corporate responsibility to respect human rights, including for companies to conduct due diligence to identify, avoid, mitigate and remedy impact on human rights for which they are responsible or indirectly involved (*246*). Although the UN Guiding Principles do not require businesses to conduct human rights impact assessments, such an assessment can help companies to meet their obligations.

Impact assessments allow identification, understanding, assessment and mitigation of the adverse effects of business projects or activities on human rights (*247*). Although such assessments are relatively new, their use has increased, including for the deployment of AI. The United Nations Special Rapporteur on Freedom of Expression noted (*3*)

> Human rights impact assessments and public consultations should be carried out during the design and deployment of new AI systems, including the deployment of existing AI systems in new global markets.

Human rights impact assessments have also been recognized in national laws as an obligation of companies. For example, the French Government enacted a law on "duty of vigilance" that requires parent companies to identify and prevent adverse impacts on human rights and the environment resulting from their activities, from the activities of companies that they control and from the activities of the subcontractors and

suppliers with which they have commercial relations (*248*). Furthermore, a EU Directive may require all companies with headquarters in Europe to conduct human rights due diligence, although the discussions will be completed only in 2021 (*249*).

Other types of impact assessment have been either proposed or implemented. One approach is an "ethical impact assessment" to identify the impacts of AI on human rights, including in vulnerable groups, labour rights, environmental rights and their ethical and social implications. A second approach, proposed by the AI Now Institute, is an "algorithmic impact assessment" for public agencies, as a "practical framework to assess automated decision systems and to ensure public accountability" (*250*). Such assessments would be both for affected communities to obtain information on how automated decision systems function and to determine whether they are acceptable and also for governments to assess how the systems are used, whether they have disparate impacts in particular on the basis of gender, race or another dimension and how to hold the systems accountable. This could be useful for governments as they turn to algorithmic decision-making for large- and small-scale health-care decisions.

Several laws have been proposed or implemented that require impact assessments, including for the use of AI for health. In 2019, two senators in the USA co-sponsored the "Algorithmic Accountability Act", which would require companies to study and adjust flawed algorithms that result in inaccurate, unfair, biased or discriminatory decisions that would affect people in the USA (*251*). It would also require companies, with enforcement by the US Federal Trade Commission, to "reasonably address" the results of such assessments, including algorithmic decisions that affect health. Such assessments would be made only for "high-risk" decisions, which would include health information or genetic data or decisions or analyses of sensitive aspects of individual lives, including their health and behaviour. The act has, however, only been proposed and is not enacted (*251*).

A separate proposal under the proposed legislation would require companies to conduct "data protection impact assessments" for high-risk information systems, such as those that store or use personal information, including health information. This would mirror the impact assessment required by law under the EU GDPR, which requires companies to conduct 'data impact assessments" of the risks of data processing operations to the "rights and freedoms of natural persons" and their impact on the protection of personal data (*252*).

Recommendations

1. Governments should enact laws and policies that require government agencies and companies to conduct impact assessments of AI technologies, which should address ethics, human rights, safety and data protection, throughout the life-cycle of an AI system.

2. Companies and developers should conduct impact assessments as per the UN Guiding Principles on Business and Human Rights, even if governments have not mandated them.

3. Impact assessments should be audited by an independent third party before and after introduction of an AI technology and published.

7.4 Research agenda for ethical use of artificial intelligence for health care

In a fast-moving field such as the use of AI for health, there are many unresolved technical and operational questions on how best to use AI. Use of AI also generates ethical quandaries. Each new application or use of AI raises opportunities and challenges that should be addressed before widespread adoption. This has been the case for the proliferation and deployment of new AI technologies during the COVID-19 pandemic.

Suggested areas of research to address emerging issues and challenges

Some ethical concerns require research to substantiate and explain the challenges. Approaches to addressing concerns should be tested and validated with research, such as on computer science or on the consequences of using AI for a particular medical need or target population. Research on each of these topics should include consideration of different countries, cultures and types of health-care systems. Pertinent research questions include the following.

- For what needs and gaps identified by health-care workers and patients could AI play a role in ensuring the delivery of equitable care?
- How is AI changing the relationships between health-care workers and patients? Do these technologies allow providers to spend more "quality" time with patients, or do they make care less humane? Do specific contextual factors improve or undermine the quality of care?
- What are the attitudes of health-care workers and patients towards the use of AI? Do they find these technologies acceptable? Do their attitudes depend on the type of intervention, the location of the intervention or current acceptance of these technologies both in the health-care system and in society?
- Has the introduction and use of AI for health exacerbated the digital divide? Or does AI, with telemedicine, reduce the gap in access to care and ensure equitable access to high-quality care, irrespective of geography and other demographic factors?
- How best can providers and programmers address any biases that will manifest in applications? What are the barriers to addressing biases?

- What method should be used to assess whether AI is more cost–effective and appropriate than existing or "low-technology" solutions in LMIC? How should governments and providers assess fair resource allocation for existing interventions and new technologies?
- Can ethical design be applied specifically to AI technologies for health?

8. LIABILITY REGIMES FOR ARTIFICIAL INTELLIGENCE FOR HEALTH

Although the performance of machine-learning algorithms is improving, there will still be errors and mistakes, for example because an algorithm has been trained with incomplete or inappropriate data, programming mistakes or security flaws. Even AI technologies designed with well-curated data and an appropriate algorithm could harm an individual. While AI technologies may be safe in practice, unforeseeable risks are likely (*253*).

Lawmakers and regulators should ensure that rules and frameworks for safety are applicable to the use of AI technologies for health care and that they are proactively integrated into the design and deployment of AI-guided technologies. Updated liability rules for the use of AI in clinical care and medicine should at least include the same standards and damages already applied to health care. It is possible that reliance on AI technologies and the risks they may pose require additional obligations and damages. This section addresses how liability regimes could evolve, approaches to compensation, specific considerations for LMIC and the role of international institutions and organizations. It does not address liability that may arise from data processing.

8.1 Liability for use of artificial intelligence in clinical care

Use of AI to support or augment clinical decision-making raises several questions. Should doctors be held at fault if they follow the suggestion of an AI technology that results in a medical error or if they ignore a suggestion that would have avoided morbidity or mortality? The answers to these questions depend largely on other choices, such as the types of behaviour encouraged or discouraged by a legal system and the standard of care as use of AI in clinical practice becomes more common.

Another choice is whether liability rules should encourage clinicians to rely upon AI to inform and confirm their clinical judgement or to deviate from their own judgement if an algorithm arrives at an unexpected conclusion. If liability rules penalize health-care providers for relying on the conclusions of an AI technology that prove to be incorrect, they may use the technology only to confirm their own judgement. While this may shield them from liability, it will discourage use of AI to its fullest potential, which is to augment and not just validate human judgement (*254*). If doctors are not penalized for relying on an AI technology, even if its suggestion runs counter to their own clinical judgement, they might be encouraged to make wider use of these technologies to improve patient care or might at least consider their use to challenge their own assumptions and conclusions.

Whether a doctor uses AI also depends on the prevailing standard of care. If AI technologies are viewed as deviating from or are not recognized as meeting the standard of care, doctors will be discouraged from using them, since, otherwise, meeting the standard of care defends (although not absolutely) medical error. If the standard of care requires use of AI technologies, physicians would essentially be mandated to integrate their use into clinical practice (*254*).

A separate but related issue is the liability of hospitals and health-care systems that select a specific technology. Hospitals could be held liable for failure to exercise due care in selecting the technology or in introducing, using or maintaining it (*115*). Generally, a hospital could be held vicariously liable for errors made by clinicians who work at the hospital. Hospitals are thus encouraged both to exercise due care in selecting technologies and to ensure that clinicians have clear guidance on how to use them for both patient care and to avoid errors that result in legal liability for the clinician and the hospital (*255*). One possibility would be to establish hospital liability by "negligent credentialing". As, generally, hospitals are liable if they do not adequately review the credentials and practice history of health workers and physicians, they could have a similar duty when introducing AI (*256*). For this, hospitals and health systems would have to have the necessary information and tools to identify appropriate AI technologies for clinical use (*256*). Hospitals should also have a duty to re-establish control of a process or system that has been automated and that now presents actual or potential risks that were not previously foreseen.

8.2 Are machine-learning algorithms products?

As AI technologies and their software are integrated into or replace medical devices, it is not clear whether they can be characterized as products. Product liability, which holds the manufacturer or developer of a technology or a good to account even if they are not at fault, is a form of strict liability in which liability is imposed even in the absence of negligence, recklessness or intent to harm (*257*).

Until now, many jurisdictions have hesitated to apply traditional product liability theory to health-care software and algorithms. Product liability could apply insofar as an algorithm is integrated in a medical device or diagnostic. Both European and US courts and new regulations regard medical software as a medical device because of its intended use (*258*). Developers may, however, escape liability because in many cases the "actual uses" of a product differ from the "intended uses", even if some of the "actual uses" could have been foreseen (*258*). Product liability may also not apply if an AI algorithm is construed as a service and not as a product.

Extension of product liability might be desirable; otherwise, patients might find difficulty in obtaining compensation (e.g. if a clinician followed the standard of care), and bringing a case to assign fault to a developer might be too costly and complex.

The design, quality assurance and deployment of AI technologies may involve many people, which could also complicate assignment of liability. Product liability could ensure that developers take all possible steps during development of an algorithm to reduce the likelihood of error, including using diverse, complete data sets to train the algorithm and improving the explainability of the software (*259*). Unforeseeable risks and safety failures could, however, limit the effectiveness of current product liability standards.

Assessment of the point to which a developer can be held strictly liable for the performance of an algorithm is complicated by the growing use of neural networks and deep learning in AI technologies, as the algorithms may perform differently over time when they are used in a clinical setting (*260*) if it is assumed that systems are allowed to update themselves and learn continuously and that use of neural networks and deep learning for AI technologies for health is acceptable and necessary.

Holding a developer accountable for any error might ensure that a patient will be compensated if the error affects them; however, such continuing liability might discourage the use of increasingly sophisticated deep-learning techniques, and AI technology might therefore provide less beneficial observations and recommendations for medical care. It could be argued that liability provisions should be written such as to discourage development of a technology that cannot be fully understood. If this were to be interpreted as requiring the explainability of the mathematical processes that allow an algorithm to learn, however, most machine-learning techniques would be banned. Liability may depend partly on how much control the developer continues to have over an AI technology. In many EU Member States, the extent of a developer's control determines whether a "development risk defence" allows the developer to avoid strict liability (*260*).

Even if developers could be held strictly liable within a product liability framework, they could avoid liability under the "learned intermediary" doctrine, which limits recovery from a manufacturer when a doctor prescribes drugs or devices (*261*) for which the manufacturer has provided adequate information, such as warnings about risks (*262*). With adequate warnings, decisions by a physician, as the "learned intermediary", break the line of causation between a product developer and the patient who has suffered harm (*262*).

8.3 Compensation for errors

A liability regime for AI might not be adequate to assign fault, as algorithms are evolving in ways that neither developers nor providers can fully control. In other areas of health care, compensation is occasionally provided without the assignment of fault or liability, such as for medical injuries resulting from adverse effects of vaccines (*263*). No-fault, no-liability compensation funds could be supplemented by requiring developers or the companies that develop or fund such technologies to

obtain insurance that would pay out for an injury or to pay into an insurance fund, with a separate fund providing compensation when an insurance pay-out is not triggered. In New Zealand, for example, patients seek compensation for medical injuries through a no-fault, no-liability scheme. Injured patients receive government-funded compensation, thereby giving up the right to seek damages, except in rare cases of reckless conduct (*264*). WHO should examine whether no-fault, no-liability compensation funds are an appropriate mechanism for providing payments to individuals who suffer medical injuries due to the use of AI technologies, including how to mobilize resources to pay any claims.

8.4 Role of regulatory agencies and pre-emption

AI technologies, like drugs and devices, will be increasingly subject to regulatory oversight and validation before use, especially as their uses expand and as clinicians increasingly rely upon them. If a commercial algorithm is approved by a regulatory agency, the doctrine of pre-emption may apply, i.e. that a decision taken by a central government agency to validate a technology will supersede any cause of action guided by civil laws (*265*). Pre-emption may not always be relevant, however, especially if the regulatory pathway for approval of an AI technology is abbreviated or regulatory approval is based on little information on how the algorithm was constructed and trained and may perform over time (*265*). Furthermore, as developers in some jurisdictions may not be held accountable for an algorithm as it evolves and learns after its sale, a doctrine of pre-emption may not be applicable if an algorithm evolves after a regulatory agency has approved the technology.

8.5 Considerations for low- and middle-income countries

Much of the literature, policy frameworks and court decisions on liability regimes are from the EU and the USA, which is where AI technologies are actively deployed. It is not known whether these approaches will be adopted in LMIC or whether those countries will take different approaches to liability. Liability rules play an important role in promoting safety and accountability, and, in some cases, they are the first and only line of defence against errors made by machine-learning technologies. Many LMIC still lack sufficient regulatory capacity to assess drugs, vaccines and devices and might be unable to accurately assess and regulate the rapidly arriving machine-learning technologies for the public good. Concern that such technologies might not operate as intended is heightened by the lack of good-quality data to train algorithms and the fact that AI technologies may have "contextual bias" (*192*). Such concern should not preclude the use of AI in LMIC, but it highlights the importance of robust, effective liability regimes. Many LMIC may wish to use AI technologies in resource-poor settings for reasons that do not apply in the EU or the USA, such as lack of health-system infrastructure.

In many LMIC, injured parties may not have access to justice, or it may be too expensive or too protracted, so that it not just difficult to obtain compensation for harm caused by AI technologies but it is also unlikely to serve as a deterrent to those responsible for the development and deployment of such technologies. Marginalized populations have even less protection and are often excluded from redress within the legal system. It might also be difficult to seek compensation if the AI technology was developed by an international company or developer with no physical presence where the harm occurs. These challenges must be addressed to increase the effectiveness of liability rules.

LMIC might have to address challenges and risks that are not often considered in high-income economies. These include lack of appropriate training data for the algorithm to ensure that it performs accurately for patients with a different physical appearance and poor connectivity, which can compromise reliable, safe use of a technology.

Even if legal systems in LMIC adopt the approaches of HIC for the introduction of AI technologies for clinical use, they will have to develop approaches that are consistent with legal practices and standards to compensate people who are harmed by such technologies, hold companies and governments accountable for the products they develop and calculate the risk–benefit for using or refusing AI technologies. WHO should work with other United Nations agencies and with governments in the design and introduction of appropriate liability rules.

Recommendations

1. International agencies (and professional societies) should ensure that their clinical guidelines keep pace with the rapid introduction of AI technologies, accounting for the evolution of AI technologies by continuous learning.

2. WHO should support national regulatory agencies in assessing AI technologies for health.

3. WHO should support countries in evaluating the liability regimes that have been introduced for the use of AI technologies for health and how such regimes should be adapted to different health-care systems and country contexts.

4. WHO and partner agencies should seek to establish international norms and legal standards to ensure national accountability to protect patients from medical errors.

9. ELEMENTS OF A FRAMEWORK FOR GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH

Human rights standards, data protection laws and ethical principles are all necessary to guide, regulate and manage the use of AI for health by developers, governments, providers and patients. Many stakeholders have called for a commonly accepted set of ethical principles for AI for health, and WHO hopes that the principles suggested in this report (See section 5.) will encourage consensus.

Use of AI for health introduces several challenges that cannot be resolved by ethical principles and existing laws and policies, in particular because the risks and opportunities of the use of AI are not yet well understood or will change over time. Furthermore, many principles, laws and standards were devised by and for HIC. LMIC will face additional challenges to introducing new AI technologies, which will require not only awareness of and adherence to ethical principles but also appropriate governance.

Governance in health covers a range of steering and rule-making functions of governments and other decision-makers, including international health agencies, for the achievement of national health policy objectives conducive to universal health coverage. Governance is also a political process that involves balancing competing influences and demands.

At the Seventy-first World Health Assembly in 2018, Member States unanimously adopted resolution WHA71.7, which calls on WHO to prepare a global strategy on digital health to support national health systems in achieving universal health coverage (*266*). A global strategy and other governance frameworks and standards established by WHO will contribute to a governance framework for AI for health. This section addresses the ethical dimensions of several areas of governance.

9.1 Governance of data

The definition of "health data" has widened dramatically over the past two decades. Successful development of an AI system for use in health care relies on high-quality data, which are used to both train and validate the algorithmic model. This section addresses the evolution of individual consent with the proliferation of health data as well as the principles, legal frameworks and measures used by governments. This section also addresses principles and mechanisms designed and used to govern health data by communities, academic or health-care institutions, companies or governments, including how these entities should share health data.

Evolving approaches to consent

As the types, quantity and applications of health data, including for commercial use, have grown, a patchwork of approaches has emerged to facilitate individuals' relation to their health data. The main challenge is safeguarding individual privacy and autonomy by controlling their data without limiting the purported benefits of their collection and use. These considerations are likely to apply whether the data are used for AI or for a relational database.

Mechanisms for individual control of data, such as informed consent, a duty of confidentiality and de-identification, may not be sufficient and may interfere with positive uses. (See section 6.3.) Therefore, several "modified" approaches to consent could be used as the quantity of health data and their possible uses increase. Consent must be given only after explanation of the consequences of providing it, including for example which data will be used and how and the consequences if consent is not given.

One form of consent that could improve individual control and choice is electronic informed consent, in which online forms and communication are used to give consent for various uses of health data (*114*). Electronic informed consent could allow users better understanding of how their data will be used and improve their control of the data. The content should, however, be presented simply so that it is readily accessible to the general public, such as with illustrations, to ensure that consent is given freely and that the risks are understood (*114*). Sage Bionetworks, for example, has established a toolkit and information guide for facilitating provision of electronic informed consent (*267*). Another approach is "dynamic consent", which allows users to modify their consent periodically for uses that they wish to permit and those that they specifically exclude (*114*). A third approach to consent, discussed below, is to seek "broad consent" from individuals to facilitate secondary use of health data without undermining their rights to privacy and autonomy.

Alternatively, governments might wish to define when consent can be waived in the public interest. This is already permissible under data protection laws if it is strictly necessary and proportionate to achievement of a legitimate aim. This implies that, in certain situations, government could have a duty to share health data for the benefit of the wider public or for other non-monetary benefits, such as better quality of life or health (*268*). Thus, consent would be waived because the data are considered a public good for which data can be "conscripted for publicly minded uses" (*128*). This could include situations in which there are clear public health benefits of using data that would otherwise be unavailable because too many individuals have opted out of sharing such data. The burden of demonstrating that lack of consent is undermining a benefit should rest with the entity that seeks to avoid consent. It could imply that obtaining health data without the specific consent of the individual is justified if the benefit is broadly distributed and outweighs violation of privacy when the risk is "low"

(*128*). A system in which benefits and risks are weighed could, however, invariably lead to sharing of data without consent, as medical benefits – whether better surveillance of disease or development of a new drug – could always be considered more important than a "low risk" of violation of privacy from use of the data.

Another concern is that a government or a company may define "public interest" in a way that is not based on public health or patient need. Whether patients share the benefits may depend on the entity with which they are shared, such as commercial actors, which may not share benefits if the medical products and services are neither affordable nor available (see below). Thus, conscripting health data with the broad goal of contributing to the public good is questionable when the data are shared with a commercial entity, whatever the intended product or service. Recent instances (described in Section 6.3) of patient data that were shared by not-for-profit entities or academic institutions with private companies without the consent of the patients have raised significant concern, as the patients were not notified that their data were shared, for what purpose or the identity of the private entity.

In Japan, an approach to resolving such conflicts was passage of the Jisedan Iryo-kiban Ho (Next Generation Medical Infrastructure Law), which permits hospitals and clinics to provide patient data to accredited private sector companies, which are responsible for making the data anonymous and searchable (*269*). Before sharing data, hospitals and clinics must inform patients and give them the right to opt out. The accredited data companies anonymize and store the data and make it available to academic researchers, pharmaceutical companies and government agencies for a fee. Accredited data companies are required to institute safeguards for cybersecurity, unauthorized use of data and unauthorized disclosure by employees (*269*).

In 2020, the EU proposed a means for use of data without consent under the concept of "data altruism", previously known as "data solidarity" (*270*). This would allow companies to collect personal and non-personal data on individuals for projects that are in the public interest. The approach seeks to limit the type of company that can collect data by specifying that it must: be constituted to meet objectives of "general interest"; operate on a not-for-profit basis and be independent of any for-profit entity; ensure that any activities related to data altruism are undertaken through a legally independent structure separate from its other functions; and can voluntarily register as a "data altruism organization" in a EU Member State. To facilitate data altruism, a common European consent form will be developed, which can be tailored for different sectors and uses.

Data altruism could raise concern. First, this form of data-sharing could lead to exceptions or "grey areas" in which health data are used for commercial purposes for which the individuals from whom the data were obtained would not wish to provide

consent. Secondly, such a regulation could be rewritten over time to redefine the entities allowed to collect data for altruistic purposes. Thirdly, even if the health data were initially used for a non-commercial objective, such as in drug discovery, the product or service that emerges might eventually be licensed to or acquired by a commercial entity rather than remaining in the public domain.

Broad consent

Several not-for-profit institutions that have deposited health data in centralized biorepositories practise principles of informed consent for sharing such data, which ensures that the person who provides data understands consent at enrolment. Any industry partner is disclosed at the time of consent, and prospective, explicit consent is given for future secondary use of the data for research (*271*). These standards do not prevent secondary use of health data, except when, for example, commercial actors that were not included in the initial consent seek to use the data or when commercial actors could otherwise gain access because they subsidize activities of not-for-profit entities that have access to the data. Even with additional standards in place, at a biorepository operated by the University of Michigan, USA, access to data was denied by a review committee for only 6 of 70 projects proposed over 2 years and only because of inadequate initial consent (*271*).

Another concern with use of health data for research arises when the data are user-generated, such as data obtained from digital devices and wearables and data supplied by users to social media and other platforms and to online patient communities. Governance of such data, which may not have been collected initially for research, is complex because of the "lack of international boundaries when using the internet" and because the "online information industry has failed to self-regulate" (*133*). Andanda suggested that one means for improving governance of such data would be to encourage health researchers to adhere voluntarily to the "Global Code of Conduct", which encourages researchers and institutions to develop context-specific codes, be fair, respectful, caring and honest when dealing with online users and practise ethically informed research practices (*133*).

A more controversial issue is creating a market or system through which individuals can buy and sell health data. Health data are sensitive personal data, linked to human agency and dignity. A system that facilitates the sale of personal data could lead to a two-tier society in which the wealthy can protect their rights and afford to limit use of their data by other parties, whereas people living in poverty may feel compelled to sell their data to access social or material benefits. A system that facilitates the sale of data would be in contravention of several human rights standards. Furthermore, while the sale of data might contribute to uses that are commercially valuable but less beneficial to individual or public health, the data market itself may not function properly and could undervalue an individual's data. The sale of data could lead to loss of control by

an individual of his or her health data. Such challenges with health data have emerged with commercial sale of blood and related products such as plasma (*272*).

Data protection

From a human rights perspective, an individual should always control his or her personal data. Individuals' right to their own data is grounded in concepts that are related to but distinct from ownership, including control, agency, privacy, autonomy and human dignity. Control may include various approaches to individual consent (see above) and also collective mechanisms to ensure that the data are used appropriately by third parties (see below). Data protection laws are rights-based approaches that include standards for the regulation of data-processing activities that both protect the rights of individuals and establish obligations for data controllers and processors, both private and public, and also include sanctions and remedies in case of actions that violate statutory rights. Data protection laws can also provide for exceptions for non-commercial uses by third parties. Over 100 countries have adopted data protection laws (*273*).

Data protection frameworks and regulations are essential for managing the use of health data. The EU GDPR, which applies to citizens and residents of the EU, irrespective of whether the data controller or processor is based in the EU, also has a global reach because it applies to non-EU citizens or residents if the data controller or processor is based in the EU. The GDPR is designed to limit the data collected about an individual to only that which is necessary, to allow collection of data only for listed legitimate purposes or with an individual's consent, and to notify individuals of data-processing activities. Health data are protected under GDPR unless an individual provides specific consent or if use of the data meets certain exceptions, such as for health-related operations or scientific research. Even when exceptions apply, data processors and controllers must respect certain obligations.

GDPR also introduced "data portability", the right of individuals to obtain their personal data in a machine-readable format from one controller that can be sent to another controller (*113*). Depending on how data portability is implemented in the EU, it could allow individuals to control their own data and to share them with additional entities. Data portability could decentralize the control and distribution of data and, with appropriate implementation, could be a novel form of data management that fosters both oversight and innovation.

Data protection regulations are enforced by data protection authorities, which develop and administer regulations, provide guidance and technical advice and conduct investigations. South Africa, which introduced a data protection regime for the first time in July 2020 with enactment of the Protection of Personal Information Act 4, will introduce enforcement in mid-2021 through several means, including administrative fines that could exceed US$ 500 000 and also civil cases and criminal liability (*274*).

Some governments have nominated additional supervisory authorities to facilitate the use of health data. The United Kingdom established a National Data Guardian in 2014 for appropriate management of health data with respect to confidentiality and to improve the use of such data for beneficial purposes. In 2018, the entity was granted the power to issue official guidance on the use of data for health and adult and social care in England (*275*).

Community control of health data – data sovereignty and data cooperatives

Measures have been taken not only to promote the individual right to privacy and autonomy over health data but also to provide discrete communities with control over their data, including health data, through the exercise of data sovereignty or creation of data cooperatives. Several indigenous communities have sought to establish control over their data through data sovereignty. Māori (the indigenous population of New Zealand) have introduced principles for data sovereignty that establish, for example, control over data, including to protect against future harm, accountability to the people who provide such data by those who collect, use and disseminate them, an obligation for such data to provide a collective benefit, and free prior and informed consent, which, when not obtainable, should be accompanied by stronger governance (*276*). Māori also recognize that the individual rights of data holders should be balanced by benefits for the community and that in some situations the collective rights of the Māori will prevail over those of individuals (*276*).

First Nations groups in Canada have also outlined principles for sovereignty over their data, with four elements: ownership of data, control of data, access to data and possession of data. It is expected that, over time, First Nation tribes will establish protocols to allow wider access to these data for uses that benefit them (*277*).

A data cooperative gives people who provide data control over their data by storing the data for the members of a cooperative. Data cooperatives allow secondary uses of such data while allowing members of the cooperative to decide collectively how the data should be used (*113*). Data cooperatives allow members to set common ethical standards, and some have developed their own tools and applications to ensure that the data are used beneficially (*113*).

Federated data

Federated data systems have grown significantly. They include collaborations between research institutions, governments and the public and private sector and within the private sector. Federated data-sharing has been defined as "a promising way to enable access to health data, including genomic data, that must remain inside a country or institution because of their sensitivity" (*278*). Data do not leave the participating organization that holds them, but authorized users can make queries that allow them to access data, for example to train an algorithm. Proponents have noted that federated data systems allow each entity to govern use of its data and

that the approach preserves privacy and security (*278*). While federated data-sharing may facilitate analysis of large data sets while maintaining local control, it does not overcome concern that informed consent might not have been sought for secondary uses of the data (*137*).

Government principles and guidelines

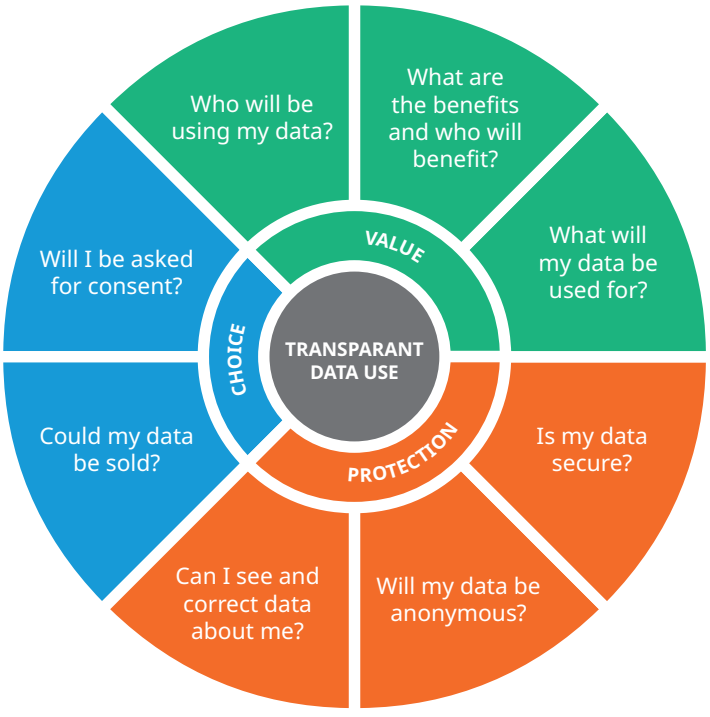
Some governments that are collecting and using health data for commercial and public sector interventions have established principles for data collection and use. The United Kingdom's NHS has established five guiding principles for a framework in which data can be used in health innovation. A notable commitment under these principles is transparency – that any commercial arrangements should be transparent, clearly communicated and not undermine public trust or confidence (*279*).

As discussed below, however, many agreements between the public and the private sector are not transparent, which raises serious concern if there are also financial conflicts of interest.

Other forms of transparency could be required, such as the transparency of sources and methods of obtaining and processing data, how and why certain types of data are excluded, the methods used to analyse the data and open discussion in publications of data bias.

In New Zealand, an independent ministerial advisory group funded and appointed by the Government conducted a wide-ranging consultation to build an "inclusive, high-trust, and high-control data-sharing ecosystem" (*280*). The guidelines include eight questions about what matters most to people in building trust in data use and whether the use of data provides value, protection and choice for an individual (Fig. 2).

Fig. 2. Elements of transparent data use



Source: reference *280*

Although the guidelines are voluntary, each entity that seeks to use the data has been asked to publish answers to these questions so that the individuals who provide the data can determine whether the values of the entity align with their preferences (*280*). WHO has introduced its own data principles (*281*), which are designed to provide a framework for data governance by WHO and to be used by staff to define the values

and standards that govern how data that flow into, across and out of WHO are collected, processed, shared and used. The five principles are as follows.

1. WHO shall treat data as a public good.
2. WHO shall uphold Member States' trust in data.
3. WHO shall support Member States' data and health information systems capacity.
4. WHO shall be a responsible data manager and steward.
5. WHO shall strive to fill public health data gaps.

WHO is also introducing a data governance framework that would introduce the necessary standards, solutions and structures to ensure the quality and integrity of WHO data, from collection, storage, analysis and validation through to use. To ensure that the principles can be put into practice, WHO will use a "hub-and-spoke" governance model to obtain feedback and approval, and data focal points at WHO will work with regional focal points on issues that arise during the ever-growing use of health data. They will also be guided by the Data Governance Committee constituted by WHO (*282*).

Data-sharing

As health data have proliferated, governments have taken steps to improve data-sharing for scientific research and also for commercial development of health AI and other health applications. In 2014, the US National Institutes of Health introduced their Genomic Data Sharing Policy, which is intended to encourage "broad and responsible sharing of genomic research data" (*283*). Legislation enacted in the USA in 2016, the 21st Century Cures Act, extended the remit and created statutory authority of the Director of the National Institutes of Health to require researchers who received awards from the Institutes to share their data and to provide the means for the Institutes to enforce data-sharing (*284*).

The Act also provides means to improve the access of individuals to their own health data, which was finalized in rules issued by the US Government in 2020 that create a requirement for health information technology providers to introduce a standards-based application programming interface to support an individual's use and control of electronic health information (*285*). Health information technology providers must meet three requirements for its interface to be certified: it must meet certain technical programming standards that ensure interoperability, it must be transparent, and it must be "pro-competitive" or promote efficient exchange, access and use of health data (*285*). The requirements for health information technology providers, such as anti-blocking or interoperability, show that governments can mandate and manage commercial use of AI and other technologies for health care.

Data hubs

Numerous data hubs pool various types of health data for use by third parties, which depend on the type of data hub. Several government-sponsored data hubs have emerged. In the USA, two such hubs are the Precision Medicine Initiative (All of Us) (*286*) and the Department of Veteran Affairs health data hub. The EU is establishing a European Health Data Space to facilitate the exchange and sharing of health data (e.g. health records, genomics, registries) for purposes such as the delivery of primary care and the development of new treatments, medicines, medical devices and services, while ensuring that people have control of their own health data (*287*).

Health Data Research UK is an independent, not-for-profit organization of 22 research institutions in the United Kingdom that collect health data and make it available to public and private entities for research on diseases and ways to prevent, treat and cure them. Principles of participation have been defined in consultation with policy-makers, the NHS, industry and the public (*288*).

Data-sharing and data partnerships with the private sector

One of the more difficult questions in the creation of government, not-for-profit or academic data hubs is how they should work with companies, either in accepting data that could improve their quality or allowing the companies to use their data for training or validation of algorithms. When commercial entities make use of such data, there is concern, which has sometimes materialized, that the people from whom they were derived did not knowingly give consent for their use for commercial purposes. There is an additional concern that such agreements are not disclosed to the public or to private sector parties to such agreements.

For example, numerous agreements signed between the Mayo Clinic, a major health system in the USA, with 16 technology companies provided the Clinic with a "revenue stream and generated crucial insights for health tech firms eager to commercialise digital products and services" (*137*). In some cases, the Clinic not only shared data with a company but subsequently took an equity stake in those companies, which provided the Clinic with additional revenue. De-identified patient data were shared without requesting consent or even notifying the people who had supplied their health data for products under development. The names of eight of the firms that signed agreements were not disclosed, and none of the contracts signed between the Mayo Clinic and its technology partners were made public (*137*).

In other cases, physicians or scientists in health-care systems who had access to raw data provided to health technology firms founded or invested in the companies. An investigation in 2018 found that board members and senior executives at the Memorial Sloan Kettering Hospital in the USA had either founded or invested in an AI start-up to improve cancer diagnosis and had used the Hospital's trove of 25 million patient tissue

slides and six decades of pathology research for the company's benefit without open bidding or transparent consideration of whether the data should be shared. Memorial Sloan Kettering had also taken an ownership stake in the company (*289*).

Some companies, either alone or in collaboration with other companies, have established health data hubs with data from one or more companies, which are used in the development of products and services. Such partnerships, which may result in useful products and services, raise concern about the transparency of the activities, oversight of activities, competition and whether such private carriers of data will seek consent or at least engage the communities and individuals that provided the data.

Recommendations

1. Governments should have clear data protection laws and regulations for the use of health data and protecting individual rights, including the right to meaningful informed consent.

2. Governments should establish independent data protection authorities with adequate power and resources to monitor and enforce the rules and regulations in data protection laws.

3. Governments should require entities that seek to use health data to be transparent about the scope of the intended use of the data.

4. Mechanisms for community oversight of data should be supported. These include data collectives and establishment of data sovereignty by indigenous communities and other marginalized groups.

5. Data hubs should meet the highest standards of informed consent if their data might be used by the private or public sector, should be transparent in their agreements with companies and should ensure that the outcomes of data collaboration provide the widest possible public benefit.

9.2 Control and benefit-sharing

The application of big data and AI for health care raises questions about how to assess and govern data control, IP and other proprietary and privacy rights that might affect the use and control of medical data and AI-driven technologies. These include asserting exclusive rights over health datasets, algorithms, software and products that include AI and the outcomes of AI-based technologies, such as medicines and diagnostic technologies. Several wider questions should be resolved, including whether health big data can or should be controlled exclusively by individuals by an appropriate form of governance or by entities that may aggregate the data. (Control of personal data is discussed above.)

A separate question is whether novel products created solely by a machine can be "owned" and, if so, whether ownership rights are conferred on the machine or on the entity that created or controls the machine. There is also the question of assigning appropriate value to the public's contribution to development of new AI technologies, such as investment in the development of algorithms, provision of data by individuals and health systems and from health data hubs accessed by private actors for the development of new AI technologies. If AI technologies are increasingly protected by exclusive rights, there is the wider question of whether they will be available, appropriate and affordable in LMIC.

Control over and benefit-sharing of big data

The central role of big data for AI, including medical big data for use of AI for health care, has led to labelling of data as the new "oil", a valuable commodity over which there will be increased commercial conflict for its control, use and access (*290*). Such labelling has been criticized as unhelpful and conceptually inaccurate (*291*, *292*). Unlike oil, the supply of data is virtually infinite, and they can be re-used in other contexts with valuable commercial or non-commercial applications. There is at least the possibility of control of and consent for use of one's data. While the intrinsic value of oil is captured once it is extracted or drilled (subject to processing and refining), data are not intrinsically valuable unless data science is used to generate something of value.

Another view is that it is not so much the commercial value of data but its use in the development and deployment of AI-based applications that is important. In this view, data are the "oxygen", an indispensable resource for the public infrastructure required for AI and data science to serve the public and private sectors (*293*). Whether data should be considered "oil" or "oxygen" (or neither) depends partly on whether exclusive rights can or should be associated with data, who should have such exclusive rights and to what extent they should impede others from access to and use of the data for public or private uses.

Several types of IP rights may apply to data and software, including protection of trade secrets, copyright, database rights (in only a few jurisdictions), regulatory exclusivity and, in rare circumstances, patent rights. Data and software as such cannot be patented in most jurisdictions, but "functional" data used in technical applications may be patented (*294*, *295*). It is beyond the scope of this publication to discuss the IP rights that could apply to large data sets or to big data, yet such rights, if they are to be expanded or minimized with respect to large data sets or big data depend on broader policy objectives and ethical considerations.

There is a conflict between sharing data and the commercial prerogatives that are protected by IP rights (*296*). On the one hand, conferring IP and related rights to health big data could discourage open sharing of the data, which is necessary to

advance scientific progress and the development of AI for health care and medicine (*93*, *295*). Public or private "owners" of health big data might not grant third parties the right to use the data to develop novel AI technologies, thereby undermining open innovation (*297*) and giving commercial entities the power to exclude competitors or engage in "rent-seeking". Questions should arise about who is allowed access, the rationale for inclusion or exclusion and the conditions under which the data will be accessible (including whether fees must be paid), especially for third parties that wish to use the data for non-commercial purposes. On the other hand, lack of IP rights to health big data could discourage some commercial investments (*297*). While the 21st Century Cures Act, enacted in the USA in 2016, encourages the sharing of data (see section 9.1), it asserts that proprietary interests supersede data-sharing interests and that the ability of the US Government to mandate data-sharing is limited by policies for prioritizing the protection of trade secrets, proprietary interests, confidential commercial information and IP rights (*284*). Similar considerations apply, for example, to the FAIR Data principles of the European Open Science Cloud, which plans to create data-sharing clouds that are "as open as possible and as closed as necessary" and does not preclude respect for IP rights or the protection of privacy rights (*298*).

An additional concern is whether sharing of health data by communities, health systems or governments in LMIC will include sharing of benefits, especially if the data are used for commercial applications of AI (*93*). If benefits are not shared, it may be either because there are no legal conventions or frameworks that mandate benefit-sharing of the uses of big data or because the entities that negotiate benefit-sharing on behalf of LMIC may have to negotiate from a weaker position (*295*). Benefit-sharing may include not only equitable access to and availability of technologies that arise from sharing health big data but also the assurance that enough investment is made in digital infrastructure, research capacity, training and infrastructure to ensure that the products of AI and big data are also generated by researchers and companies in LMIC (*295*). New technologies that require "state-of-the-art" capacity, such as quantum computing, might exacerbate inadequate benefit-sharing.

Thus, while IP rights could be adjusted case by case to encourage open innovation, investment or benefit-sharing, control (and IP rights to assign control) may be inappropriate to encourage widespread use and application of health data, in view of numerous competing considerations, including an individual's right to privacy and control (*299*), society's interest in scientific progress and the development of AI-guided technologies, commercial interest in exploiting such data for profitable activities and the interest of data contributors (communities, health systems, governments) in sharing the benefits generated by third parties (*299*).

It has been recommended that the focus be not on recalibrating or introducing new IP rights, which could impede data-sharing or intensify competing claims to control

of data, but instead on establishing a legal framework based on custodianship (*93*). Custodianship, or responsible oversight with ethical values, can ensure access to data, promote fair data-sharing and preserve privacy. While those who provide data maintain limited control, certain decisions are delegated to data custodians with custodial rights – and not control (or IP rights) – over big data. Custodial rights can include protecting the privacy of those who contribute data, disseminating research findings, ensuring freedom of scientific enquiry and providing attribution to those who invest in creating databases and agreeing on terms of use and access (*295*).

Ownership of AI-based products, services and methods

Products and services created with AI and big data could be patented or subject to other IP rights. These include algorithmic models that can be used in drug discovery and development and the end-products of such uses of AI, such as new medicines, medical devices or diagnostic methods. Thus, as noted in section 3.2, the announcement by DeepMind of a new AI model, AlphaFold, may result in real progress in the development of new medicines but might be heavily protected by patents and other forms of IP and therefore not widely available. If other AI technologies and tools that could accelerate drug development are not placed in the public domain (e.g. without IP protection) and are not available for licensing on a royalty-free basis or under reasonable terms and conditions, the companies that own such technologies will exert greater power and control over the development of new medical technologies and services.

An overlying concern in patenting (and other forms of ownership) of AI-generated inventions is therefore that IP rights could exclude affordable access to the products or services and that patent holders engage in rent-seeking behaviour to recuperate investments and earn outsized profits. As novel medicines, diagnostic methods and other products and services developed with AI may depend on publicly generated health data and other public-sector investments in AI and health-care infrastructure for identification, testing and validation, the question arises of whether the public investment will be rewarded, including by ensuring affordable access to the product. All science, including advances in AI, has been based on decades of publicly funded academic research.

Assessing ownership is especially difficult when a product or research output is the result of a PPP for which governments may have provided funding and other forms of support but which maintain limited or no ownership of the research output. Ensuring a role for government in both the development of new AI technologies and the ownership of the outcomes would be fairer for the governments and citizens that contribute resources and data to collaboration with the private sector.
Another concern is that issuing time-limited patent monopolies for such inventions, even if they encourage innovation, may discourage the companies that own AI

technologies from considering the needs of people living in poverty in LMIC when developing or adapting such products. Thus, as AI is used more frequently to develop new technologies to improve health care, including new medicines, the use of incentives outside the patent system, such as those that separate the cost of research and development from the expectation of high prices, could encourage companies that develop these technologies to invest in use of AI or to adapt new products to meet global public health needs.

Companies might refuse to disclose data that they consider an "essential facility" for developing, for example, a much-needed vaccine or choose to collaborate only in strategic areas of data application and with control of the data that are shared, with whom and under which conditions. This could replace healthy competition by collusion, with future effects on competition that are difficult to assess. Antitrust (competition) authorities will have to consider new approaches to address such issues (*297*).

Several legal issues will affect the patenting of AI technologies. One is whether AI-guided machines that develop new products or services can be considered inventors, which would lead to questions about defining the threshold for meeting the criteria for patenting an invention, such as an inventive step. Some legal experts have argued that recognition of machines as inventors would encourage the development of creative, powerful machines that can generate new innovations (*300*). If, however, most such machines are owned by a few companies, the benefits of the inventions will accrue to those few companies, which will wield significant power through exclusive rights and use the machines to capture an entire field of technology. In January 2020, the European Patent Office ruled that machines cannot be listed as inventors under current patent laws (*301*), and the US Patent and Trademark Office has issued a similar decision (*302*).

Another legal issue is whether diagnostic methods and algorithms can be patented. While in the USA securing patent protection for diagnostic methods and mathematical models is highly restricted, the EU has provided several grounds for the issuance of patents (*303*). While patent monopolies could encourage the development of new technologies with greater medical benefits, patenting of such methods and services could limit their diffusion, access and benefit-sharing with the populations that contributed the data used to train or validate the technology.

Recommendations

1. WHO should ensure clear understanding of which types of rights will apply to the use of health data and the ownership, control, sharing and use of algorithms and AI technologies for health.

2. Governments, research institutions and universities involved in the development of AI technologies should maintain an ownership interest in the outcomes so that

the benefits are shared and are widely available and accessible, particularly to populations that contributed their data for AI development.

3. Governments should consider alternative "push-and-pull" incentives instead of IP rights, such as prizes or end-to-end push funding, to stimulate appropriate research and development.

4. Transparency in regulatory procedures and in interoperability should be enhanced and should be fostered by governments as deemed appropriate.

9.3 Governance of the private sector

The private sector plays a central role in the development and delivery of AI for health care. The "private sector" ranges from small start-ups to the world's largest technology companies, as well as companies that provide many of the materials necessary for AI, including health data collected by companies that supply wearable devices, data aggregators and software firms that write new algorithms for use in health care. Furthermore, many companies that were already providing products and services are transforming their businesses to integrate AI and big data. These include biopharmaceutical companies, diagnostic and medical device firms, insurance companies, private hospitals and health-care providers. Companies that are developing AI technologies for use in health care are also providing these applications and services outside the health-care system, raising the question of how such health-care provision should be regulated.

This section addresses several issues related to the governance of such companies: To what extent should oversight and governance of the private sector be enforced by companies collectively or individually? What challenges and opportunities for effective governance are associated with PPPs for AI for health care? What are the challenges of oversight and governance of large technology companies involved in the use of AI for health? How should governments manage the growth of health-care services provided by companies outside the health system? How can governments ensure that they are effectively overseeing the private sector?

The role of self-governance

As companies often push the boundaries of innovation and act much more quickly than can be anticipated by regulators, governments and civil society, they often first set the rules in the code that they write, the services they design and the corporate practices and terms of services they offer (*304*). As some innovations have raised concern, companies have strengthened their internal processes and measures to avoid criticism and have pursued collaborations and partnerships. Thus, some have introduced their own ethical principles and internal processes for integrating ethical considerations into their business operations (*156*). This includes integrating ethics

into the design of new technologies and design-related approaches to privacy and safety. Companies have also launched multi-stakeholder initiatives to develop best practices (*305*), although there is no such initiative yet for the use of AI for health. While integration of ethics into a company's operations is welcome, it raises as many concerns as hopes, the concerns including that companies may be engaging in "ethics-washing" and that the measures are intended to forestall regulation instead of adapting to oversight (*156*). In some companies, efforts by ethics teams to address ethical challenges and concerns may be discouraged or have repercussions. For example, a news report stated that Google had fired an AI ethics researcher who criticized Google's "approach to minority hiring and the biases built into today's artificial intelligence systems" (*306*). Even if attempts to formulate and integrate ethics into daily company operations are taken seriously, other challenges may limit their effectiveness.

First, the incentives and values of AI firms and developers may differ from those of the patients, health-care providers and health-care systems (*306*) that will use such products and services but have no role in establishing the culture or norms in which the products and services are developed (*307*). For example, large technology companies, which are based in only a few countries, may adopt values and belief systems that are not appropriate for other countries, health-care systems or communities. More generally, while medicine is guided by the objective of promoting the health and well-being of patients, an AI developer who is developing a product or service that provides benefits is ultimately working in the interests of the company to develop a profitable service or product and, in the case of publicly traded companies, for their shareholders (*305*). While medical professionals have a long-standing fiduciary relationship with patients, AI developers, however well-intentioned and with emerging expectations and legal obligations to protect individual privacy, have no fiduciary duty to patients or health-care providers. This complicates any attempt by an individual or a company to put the health and well-being of patients first (*305*).

Secondly, the ethical norms adopted by companies might be difficult to translate into practice (*156*), either because AI developers have no suitable methods of doing so, as AI is a relatively new technology, or practical measures to adhere to high-level ethical norms may be difficult to reconcile with a culture of fast growth, fast failures and getting first to the market. Ethical principles may therefore be "watered down", modified or rendered ineffective. It may also be difficult to determine whether ethical norms are written into the source code for an AI technology, whereas, in the practice of medicine, numerous structures built over time, including professional societies and boards, ethics review committees, accreditation and licensing schemes, peer self-governance and codes of conduct, determine and shape what is acceptable, and bad practices and bad actors can be identified quickly (*305*).

Thirdly, there are insufficient legal and professional accountability mechanisms to reinforce good-faith efforts of firms to turn ethical principles into practice (*305*). Unlike the medical profession, AI developers and technology firms have no effective self-governance mechanisms and do not face the legal penalties and repercussions of other professions, especially the medical profession. Accountability mechanisms in the medical profession reinforce its fiduciary duty to patients and are reinforced by sanctions to deter poor practices. AI development does not include professional or legally endorsed accountability mechanisms (*305*).

Fourthly, it is questionable whether companies can govern their own AI products and services effectively to minimize any harmful direct or indirect impact on health care. For example, social media companies such as Facebook play an important role in sharing health information through platforms such as Facebook and WhatsApp. There has recently been significant concern about the spread of misinformation and disinformation on its platforms that undermines medical and public health information issued by governments and international agencies, and this has increased during the COVID-19 pandemic. The company has taken steps to address misinformation and disinformation, including a partnership with WHO to create a chatbot on Facebook Messenger and WhatsApp to provide accurate information through the WHO Global Alert Platform (*308*).

A study by a not-for-profit group, Avaaz, found, however, that the spread of medical disinformation and misinformation on Facebook far exceeded information from trustworthy sources such as WHO. The most popular "super spreader" sites received four times more clicks than bodies such as WHO and the US Centers for Disease Control and Prevention (*309*). According to Avaaz, this was due largely to amplification of public pages that featured misinformation in Facebook's algorithm. During the early stages of the COVID-19 pandemic, in April 2020, "disinformation sites attracted an estimated 420 million clicks to pages peddling harmful information – such as supposed cures for SARS-CoV2" (*310*). Only 16% of misleading or false articles displayed a warning label by Facebook third-party fact-checkers (*310*). Furthermore, while Facebook has subsequently sought to address misinformation on COVID-19 by deleting false posts and directing users to valid information (*311*), some researchers have criticized Facebook for not identifying the misinformation and correcting it (*312*).

The concern that a few companies manage information critical to the public good extends to whether such companies might withhold such information because of public policy or corporate disputes. In 2021, Facebook, having been unable to reach an agreement with the Australian Government about a new law that would require the company to pay news publishers for the content it placed on its site, decided to block users from accessing news stories on its platform (*313*). The block included access to Australian state government health websites and prevented the state governments

from posting on the website, even as the Government was preparing public announcements about vaccination against COVID-19 (*314*). Websites that posted misinformation about vaccines were unaffected (*315*).
None of these concerns should be a reason for companies not to invest in improving the design, oversight and self-regulation of their products. The improvements could include licensing requirements for developers of "high-risk" AI, such as that used in health care, which would bring AI developers in line with requirements in the medical profession and increase trust in their products and services. International standards organizations have made important contributions to improving applications of health information technology, from data structure and syntax to privacy and implementation. For instance, the International Standardization Organization (*316*), Health Level Seven International (*317*) and other organizations have contributed to the governance of information technology, including machine learning, and such standards have been described as carrying ethical weight (*177*).

Public–private partnerships for AI for health care

PPPs are common in health care, and, unsurprisingly, PPPs are emerging in the field of AI for health care. In one type of PPP, raw data are provided by the public sector, such as electronic medical records and other health data collected in health-care systems and hospitals, and these are used by one or more companies to develop products and services, such as diagnostic methods and predictive algorithms.

Supporters of PPPs in both government and industry emphasize the benefit of leveraging the resources and innovative capacity of companies to generate products and services. Presumably, in such collaborations, governments can oversee the activities of the private companies and safeguard the public interest. There are, however, challenges in ensuring effective governance of the private sector. First, there is a significant asymmetry in information and skills between companies and government agencies in such partnerships. Companies often hire trained professionals who are well versed in the technology in question and in the parameters of a negotiated partnership. A second challenge is that the "social license" granted to the public sector for use of certain resources, such as patient data, may not extend to private companies, which may not be trusted and have goals and objectives that may not be aligned with public expectations (*216*). Thirdly, public sector entities have several competing priorities that may undermine a government's ability to oversee the partnership effectively. A public sector entity may have difficulty in reconciling the objective of successful development of a new product or service, the obligation to protect the rights of individuals and patients and the wider responsibility to regulate all the operations of a private sector partner effectively.

Fourthly, there is often concern that the contributions of the public sector and the community (technology, data, funding, expertise, testing sites) are not considered when allocating ownership rights (if any) to a technology between the public and private sector and in setting the price of such technologies or the rules under which the technology is used (*216*). If the public sector and communities make significant contributions to a partnership but are not full beneficiaries, such collaborations may be considered exploitative.

Governance and oversight of large technology companies

Large technology companies, especially those located in China and the USA, are expected to play a central role in the development and deployment of AI for health, through partnerships, in-house development of AI or acquisition of other companies. The role and involvement of these companies raises further considerations for oversight of the private sector. Large technology companies, of which there are only a few, wield significant power in the field of AI because of their human, economic and technical resources, the data accumulated from their products and services, the political influence they may be able to exert through their relationships and partnerships with governments and their staff (see below) and their ability to use their platforms to introduce products and services to large numbers of users, who are regularly connected to their platforms.

Over time, large technology companies may develop even more diversified products and services. Google is developing a range of diagnostic applications that are still being examined for safety and efficacy, and its parent holding company, Alphabet, has launched a new health insurance service that will work in partnership with SwissRe (*318*).

Companies may also launch products and services that could compete with, replace or introduce a function or process that is usually managed by a government. Tencent has introduced an application that uses information voluntarily supplied by individuals to determine the type of health-care provider a patient should consult, partly to resolve a practice in China whereby patients use their own research or intuition to seek medical advice from specialists in areas unrelated to their condition.[8] The growth of telemedicine is providing opportunities for company-owned platforms to move patients to their platforms, and they are enrolling doctors to provide services via the platform. For example, Tencent WeDoctor, which works with the Government, has enrolled at least 240 000 providers onto its platform and also 2700 hospitals and 15 000 pharmacies. At least 27 million monthly users consult the "health-care collaboration platform" for an AI-guided or a remote consultation. Users are then matched with the appropriate specialist in the health-care system (*319*). This could mean that, in the long term, governments might not so much regulate companies that provide such services but might depend on them to fill gaps and manage parts

[8] Presentation by Alexander Ng, Tencent, 27 August 2020, to the WHO Expert Group on AI for health.

of the health-care system. Technology companies may supply the infrastructure for operation of health-care services, which also creates dependence of governments on the services and capabilities of the companies, rather than regulating the industry to serve the needs of the government and the public.
As noted above, technology companies have begun to issue guiding principles for the use of AI; however, they are sometimes viewed as "ethics washing", may create a gap in responsibility (assigning responsibility for retrospective harm), do not involve the public in their development and may be administered in a way that is not transparent to the public or to governments, with no involvement of the public or an independent authority for oversight of adherence to the principles.

Provision of health care by the private sector outside the health-care system
The proliferation of AI applications for health outside the health-care system may extend access to some health-care advice; however, such applications raise new questions and concerns. An application may be developed without appropriate reference to clinical standards; it may not be user friendly, especially for follow-up services or procedures; patient safety may be compromised if individuals are not connected to health-care services, such as lack of assistance to individuals with suicidal ideation who use an AI chatbot; the efficacy of applications such as chatbots that may not have been tested properly may be inadequate; and applications may not meet the standards of privacy required for sensitive health data (*319*). As such applications are not necessarily labelled as health-care services and may not even be known to governments, the overall quality of health care could be compromised, and people with no other options may be relegated to subpar services. Governments should identify these applications, set common standards and regulations (or even prevent some applications from being deployed to the public) and ensure that individuals who use the applications retain access to appropriate health-care services that cannot be provided online.

An enabling environment for effective governance of the private sector
Appropriate governance of the private sector must overcome a number of hurdles. One is the power of many of the companies involved in delivering AI for health care. Many of them employ former government officials and regulators, who are asked to lobby and influence policy-makers and regulators charged with overseeing the use of AI for health care. This can affect the ability of governments to act independently of companies.

A second challenge is that many of the technologies developed by companies are increasingly difficult to evaluate and oversee, partly because of their growing complexity, including the use of black-box algorithms and deep learning methods. The growing complexity has encouraged both governments and companies to consider models of "co-regulation", whereby each party relies on the other to assess and regulate a technology. While such models of oversight may assist governments in

understanding a technology, they may limit the government's exercise of independent judgement and encourage them to trust that companies are willing to strictly self-regulate their practices.

Improving governance of the private sector in other ways will require more independent in-house expertise and information so that governments can evaluate and regulate company practices effectively. Thus, capacity-building of government regulators and transparency will both play roles in improving government oversight of the private sector. Such measures could include greater transparency of the data collected and used by private companies, how ethical and legal principles are integrated into company operations and how products and services perform in practice, including how algorithms change over time.

Recommendations

1. Governments should ensure that the growing provision of health-related services through online platforms that are not associated with the formal health-care system is identified, regulated (including standards of privacy protection guaranteed within health-care systems) and avoided for areas of health care in which the safety and care of patients cannot be guaranteed. Governments should ensure that patients who use such services also have access to appropriate formal health-care services when required.

2. Governments should consider adopting models of co-regulation with the private sector to understand an AI technology, without limiting independent regulatory oversight. Governments should also consider building their internal capacity to effectively regulate companies that deploy AI technologies and improve the transparency of a company's relevant operations.

3. Governments should consider establishing dedicated teams to conduct objective peer reviews of software and system implementation by examining safety and quality or general system functionality (fitness for purpose) without requiring review or approval of a code.

4. Governments should consider which aspects of health-care delivery, financing, services and access could be supplied by companies, how to hold them accountable and which aspects should remain the obligation of governments.

5. Public-Private Partnerships (PPPs) that develop or deploy AI technologies for health should be transparent (including in the terms and conditions of any agreement between a government and a company) through meaningful engagement by the public. Such partnerships should prioritize protection of individual and community rights and governments should seek ownership rights to products and services so that the outcomes of the PPP are affordable and available to all.

6. Companies must adhere to national and international laws and regulations on the development, commercialization and use of AI for health systems, including legally enforceable human rights and ethical obligations, data protection laws, measures to ensure appropriate informed consent and privacy.

7. Companies should invest in measures to improve the design, oversight, reliability and self-regulation of their products. Companies should also consider licensing or certification requirements for developers of "high-risk" AI, including AI for health.

8. Companies should ensure the greatest possible transparency in their internal policies and practices that implicate their legal, ethical and human rights obligations as established under the UN Guiding Principles on Business and Human Rights. They should be transparent about how those ethical principles are implemented in practice, including the outcomes of any actions taken to address violations of such principles.

9.4 Governance of the public sector

Use of AI in the public sector has increased recently, although it lags behind adoption by the private sector. In 2019, OECD identified 50 countries that have launched or are planning to launch national AI strategies, of which 36 plan to or have issued separate strategies for public sector AI (*320*). In 2017, the United Arab Emirates was the first country in the world to have a designated minister for AI, which has resulted in increased use of AI in the health-care system, such as "pods" to detect early signs of illness, AI-enabled telemedicine and use of AI to detect diabetic retinopathy (*321*). Although use of AI has increased in the public sector, a review of nearly 1700 studies found only 59 on use of AI in the public sector (*320*). There is no comprehensive account of how governments are advancing the use of AI or integrating it into health care. The OECD identified six broad roles for governments in AI, as a:

- financier or direct investor in AI technologies in both the public and the private sector;
- "smart buyer" and co-developer, including PPPs and other forms of collaboration with companies;
- regulator or rule-maker;
- convenor and standard setter;
- data steward; and
- user and services provider.

This section briefly addresses how governments should use AI ethically as investors in AI technologies, as smart buyers and/or co-developers and as users and service providers. It also addresses concern about ethics and human rights with increased use of AI to manage social protection and welfare, programmes that often directly influence access to health-care services and indirectly affect human health and well-being.

Assessing whether AI is necessary and appropriate for use in the public sector

As for any use of AI by health professionals, governments must assess whether an AI technology is necessary and appropriate for the intended use and can be used according to its laws. The assessment could include an evaluation of whether use of AI is appropriate. In India, the Government's internal think tank, Niti Aayog, has proposed constitution of an ethics committee to review procurement of AI in the public sector. According to a draft proposal released in 2020, the committee "may be constituted for the procurement, development, operations phase of AI systems and be made accountable for adherence to the Responsible AI principles" (*322*). A requirement that both ministries of health and public and private health-care providers observe legal and ethical standards in the procurement of AI can encourage appropriate design of AI technologies and provide a safeguard against harm.

The Government of the United Kingdom has established an analytical framework for use of AI (*323*), which consists of the following: whether the available data contain the required information; if it is ethical and safe to use the data and consistent with the Government's data ethics framework; if there are sufficient data for training AI; whether the task is too large or repetitive for a human to undertake without difficulty; and whether AI will provide information that a team could use to achieve real-world outcomes.

Accountability through transparency and participation

Governments are increasingly required to disclose the use of algorithms in services and operations in order to promote accountability for the use of AI, and many data protection laws require that decisions not be taken solely by automated systems and that use of automated decision-making be prevented in certain contexts. In France, the Government is required to provide a general explanation of how any algorithm it uses functions, personalized explanations of decisions issued by algorithms, justification for decisions and publication of the source code and other documentation about the algorithms (*320*).

In general, there is growing expectation that governments will be transparent about their use of AI, including whether they are investing in AI, engaged in partnerships with companies or developing AI independently in state-owned enterprises or government agencies. It is also expected that governments will be transparent about any harm caused by use of AI and the measures taken to redress any harm. A review conducted by the United Kingdom Committee on Standards in Public Life found that the British Government (during the period examined) had not met established principles of openness and noted that "under the principle of openness, a current lack of information about government use of AI risks undermining transparency" (*324*).

Yet, transparency may not be sufficient to ensure that government use of algorithms will not result in undue harm, especially for marginalized communities and populations. Greater public participation by a wide range of stakeholders is necessary to ensure that decisions about the introduction of an AI system in health care and elsewhere are not taken only by civil servants and companies but are based on public participation of a wider range of stakeholders, including representatives of public interest groups and leaders of vulnerable groups that are often not involved in making such decisions. Their perspectives should be obtained before and not only after identification of an adverse effect, which is too late.

Appropriate collection, stewardship and use of data

The collection, storage and use of data according to ethical and legal standards also applies to governments. Government use of data is prone to abuse, whether through the sale or provision of data to private companies that violates the public trust or sharing data obtained or collected for health-care purposes in other government programmes, including enforcement of immigration laws or criminal justice. Such health data, which often include information on location or behaviour, can then be used to infringe on civil liberties directly. These uses of data undermine trust in the health-care system and the willingness of individuals to provide data and use AI technologies that are intended to improve the administration of health care and medicine.

Governments also face risks of bias in data that are collected for the development of AI for use in the public sector. The obligation of the public sector to remain objective may be undermined, as the "prevalence of data bias risks embedding and amplifying discrimination in everyday public sector practice" (*325*). The review of use of AI in the public sector in the United Kingdom also found that "data bias is an issue of serious concern, and further work is needed on measuring and mitigating the impact of bias" (*324*).

Risks and opportunities in use of AI for provision of public services and social protection

Governments have used AI to provide public services, including assessment of whether an individual qualifies for certain services, in what is known generally as the "digital welfare state". Thus, digital data and technologies are used to automate, predict, identify or disqualify potential recipients of social welfare. While some have championed this use of AI as a means of eliminating redundant and repetitive tasks that both saves resources and gives government employees more time to address more difficult issues (*325*), there is concern that the digital welfare state could undermine access to social services and welfare and especially affect poor and marginalized populations. According to a report by the United Nations Special Rapporteur on extreme poverty and human rights, the digital welfare state could become a "digital dystopia", constricting budgets intended for the provision of services, limiting those who qualify for government services, creating new conditionality and introducing new sanctions to discourage the use of

services (*326*). The report also notes that administering a welfare state through a digital ecosystem can exacerbate inequality, as many poor and marginalized individuals do not have adequate access to online services (*326*). Although the report does not discuss use of AI to provide or refuse health-care services, such use could affect the provision of health care in the public sector or, for example, the provision of health insurance through the public or private sector.

Recommendations

1. Governments should conduct transparent, inclusive impact assessments before selecting or using any AI technology for the health sector and regularly during deployment and use. This should consist of ethics, human rights, safety, and data protection impact assessments. Governments should also define legal and ethical standards for procurement of AI technologies and require public and private health-care providers to integrate those standards into their procurement practices.

2. Governments should be transparent about the use of AI for health, including investment in use, partnerships with companies and development of AI in state-owned enterprises or government agencies, and should also be transparent about any harm caused by use of AI.

3. Governments and national health authorities should ensure that decisions about introducing an AI system for health care and other purposes are taken not only by civil servants and companies but with the democratic participation of a wide range of stakeholders and in response to needs identified by the public health sector and patients. They should include representatives of public interest groups and leaders of marginalized groups, who are often not considered in making such decisions.

4. Governments should develop and implement ethical, legally compliant principles for the collection, storage and use of data in the health sector that are consistent with internationally recognized data protection principles. In particular, governments should take steps to avoid risks of bias in data that are collected and used for development and deployment of AI in the public sector.

5. Governments should ensure that any use of AI to facilitate access to health care is inclusive, such that uses of AI do not exacerbate existing health and social inequities or create new ones.

9.5 Regulatory considerations

The largest national regulatory agencies, such as the Food and Drug Administration in the USA, have been developing guidance and protocols to ensure the safety and efficacy of new AI technologies; however, other regulatory agencies may have neither

the capacity nor the expertise to approve use of such devices. A WHO working group has been formed to address regulatory considerations for the use of AI for health care and drug development and will issue a report and recommendations in 2021. The present guidance identifies several ethical concerns that could be addressed by regulatory agencies and the challenges that could arise.

Does regulation stifle innovation?

It is commonly asserted that stringent regulations will limit innovation and deprive health-care systems, providers and patients of beneficial innovations. A balance must be struck between protecting the public and promoting growth and innovation (*159*). Use of AI for health is still new and often untested, and policy-makers and regulators must consider numerous ethical, legal and human rights issues. For example, regulators must identify those applications and AI-based devices that may be best described as "snake oil", a euphemism for deceptive marketing, health-care fraud or a scam, which either misrepresents what an application can do, provides misinformation or persuades vulnerable individuals to follow health advice that may be contrary to their well-being (*327*).

Applications that provide no therapeutic or health benefit might be introduced solely for collecting health and biological data for use in commercial marketing or to encourage patients to pay for irrelevant or unproven health interventions (*328*). For example, an academic obtained data from 300 000 Facebook users who were told that the data were for a "psychological test". Their data and data from an estimated 50 million other users linked to them (Facebook "friends") were then sold to Cambridge Analytica, which used them to build a software program to predict and influence choices at the ballot box (*329*). Such malicious use of data collected nominally for academic or health purposes could expose health systems, health providers and companies that provide health-related AI services to significant risk.

Regulation could differ according to risk, such that those who are especially vulnerable, including people with mental illness, children and the elderly, are protected from misinformation and bad advice from health applications that exploit rather than assist such individuals (*159*). People living in resource-poor settings, in countries with inadequate resources to regulate and monitor adverse consequences of AI applications and with diseases that result in marginalization and discrimination, such as HIV/AIDS or tuberculosis, also require greater protection and oversight by regulatory agencies than users of applications for lifestyle or wellness.

Transparency and explainability of AI-based devices

The black box of machine learning creates challenges for regulators, who may be unable to fully assess new AI technologies because the standard measures used to assess the safety and efficacy of medical technologies and scientific understanding

and clinical trials are not appropriate for black-box medicine (*255*). Complex algorithms are difficult for regulators to understand (partly because of lack of expertise in regulatory agencies) and difficult for developers to explain.

Improving the scientific understanding (explainability) of an algorithm is considered necessary to ensure that regulators (and clinicians and patients) understand how a system arrives at a decision. Explainability is also a requirement of the EU's GDPR and is being introduced into legislation in other countries experiencing proliferation of AI for health care and other fields (*116*). It has been argued that, if a trade-off is to be made between transparency and accuracy, transparency should predominate. This requirement may, however, not be possible or even desirable in the medical context. While it is often possible to explain why a specific treatment is the best option for a specific condition, it is not always possible to explain how that treatment works or its mechanism of action, because medical interventions are sometimes used before their mode of action is understood.

Trust in decisions and expert recommendations depends on the ability of experts to explain why a certain system is the best option for achieving a clinical goal. Such explanations should be based on reliable evidence of the superior accuracy and precision of an AI system over alternatives. The evidence should be generated by prospective testing of the system in randomized trials and not their performance against existing datasets in a laboratory.

Understanding how a system arrives at judgements may be valuable for a variety of reasons, but it should not take precedence over or replace sound, prospective evidence of the system's performance in prospective clinical trials. Explanations of how a system arrives at a particular decision could encourage use of machine-learning systems for purposes for which they are not well suited, as the models created by such systems are based on associations among a wide range of variables, which are not necessarily causal. If the associations are causal, practitioners might rely on them to make decisions for which the system has not been tested or validated. Requiring every clinical AI decision to be "explainable" could also limit the capacity of AI developers to use AI technologies that outperform older systems but which are not explainable (*116*).

Clinical trials provide assurance that unanticipated hazards and consequences of AI-based applications can be identified, addressed and avoided entirely, and additional testing and monitoring of an approved AI device can demonstrate its performance and any changes that may occur after it has been approved. Clinical trials, especially those carried out with diverse populations, can also indicate whether an AI technology is biased against certain sub-groups, races or ethnicities (see below). Clinical trials may not, however, be appropriate because of their cost, because it takes a long time to conduct a trial properly, because the validity of the results may be called into question

if an algorithm is expected to change over time with new data, and because AI-based technologies and products are increasingly personalized to smaller populations and therefore more difficult to test with enough individuals (*255*).

Clinical trial designs and statistical analysis strategies should be re-evaluated, and innovation should be encouraged in these areas of AI validation. While AI should properly be validated in clinical trials or other applicable ways, AI itself could potentially allow even more accurate trials of device or drug effectiveness with smaller patient populations through enhanced patient–trial matching, data analytics efficiency and other approaches. This might become relevant during the COVID-19 pandemic as recruitment and access to health-care facilities is challenged.

Regulators could introduce "lighter premarket scrutiny" in the place of clinical trials for AI technologies for health, by assessing the safeguards put in place by developers, the quality of the data used, development techniques, validation procedures and "robust post-market oversight". This might, however, be difficult to implement in practice, especially post-market oversight of novel algorithms (*255*), and may be too late to prevent harm to people who are especially vulnerable, such as those who have no access to a health-care provider who could protect them from a misguided diagnosis or advice. The transparency of the initial dataset could be improved, including the provenance of the data and how they were processed, as could the transparency of the system architecture (*115*). Such transparency would allow others to validate an AI technology independently and increase the trust of users.

While greater transparency of the components of an AI system, including its source code, data inputs and analytical approach, can facilitate regulatory oversight, some transparency may misplace focus. Reviewing lines of code would be time-consuming and unlikely to be informative in comparison with the performance, functionality and accuracy of the system both before and after it is integrated into a health-care system.

Addressing bias

Regulatory agencies should create incentives to encourage developers to identify and avoid biases. One example is the addition of measures to a precertification programme hosted by the US Food and Drug Administration, the Digital Health Innovation Action Plan (*330*). The programme already assesses medical software on the basis of criteria of excellence, including quality. The criteria for quality and other criteria set by regulatory agencies could include the risk of bias in training data (*330*). Robust post-marketing surveillance to identify biases in machine-learning algorithms, including in collaboration with providers and communities likely to be affected by biased algorithms, could improve regulatory oversight.

Ethical considerations for LMIC and HIC with poor health outcomes

LMIC often have insufficient regulatory capacity, so that they are unable to assess the safety and efficacy of new technologies. Regulatory agencies in LMIC could consider either relying on regulatory approval of AI technologies in HIC or use of collaborative registration procedures to ensure that new technologies are appropriate for use. Global harmonization of regulatory standards would ensure that all countries benefit from rigorous testing, transparent communication of outcomes and monitoring of a technology's performance. International harmonization of regulatory standards, based on those of HIC, or reliance on other regulatory agencies or the assurances of product developers is founded on the assumption that the criteria used to develop or assess a new technology in HIC is appropriate for LMIC contexts and populations. This may not be the case, and it is likely that AI health technologies cannot be transposed between divergent settings, including between LMIC and HIC (*115*). This may be due not only to the types of data used to train the algorithm but also to the assumptions and definitions used in developing an AI technology, such as what constitutes "healthy", which may be defined by a small group of developers located in one company or country and validated by regulators in HIC with no consideration of whether the assumptions are appropriate for LMIC (*183*).

Regulators may also make assumptions about the context in which an AI technology was introduced. AI technologies may have "contextual bias", whereby the algorithms may not recommend safe, appropriate or cost–effective treatments for low-income or low-resource settings (*193*) or for countries that have resources but in which segments of the population still have poor health outcomes, as is often the case in some HIC. The developer of a technology for a high-income setting in which most of the population have good health outcomes may neither anticipate nor build an AI technology to anticipate differences from LMIC settings or from other HIC with poor health outcomes, and a regulator, even if it requires prospective clinical trials, may not require data on how the technology operates in LMIC or certain high-income settings.

While the transparency of the data used to train algorithms, the context in which an algorithm is trained and other material assumptions are necessary, they may only delay use of an AI technology, thus avoiding harm, but not bestow any benefit. Improving the performance and use of AI technologies in LMIC and certain HIC and ensuring that the technologies are adapted to reality will require different incentives, approaches and developers of technologies that are appropriate for all people (*193*).

Recommendations

1. Governments should introduce and enforce regulatory standards for new AI technologies to promote responsible innovation and to avoid the use of harmful, insecure or dangerous AI technologies for health.

2. Government regulators should require the transparency of certain aspects of an AI technology, while accounting for proprietary rights, to improve oversight and assurance of safety and efficacy. This may include an AI technology's source code, data inputs and analytical approach.

3. Government regulators should require that an AI system's performance be tested and sound evidence obtained from prospective testing in randomized trials and not merely from comparison of the system with existing datasets in a laboratory.

4. Government regulators should provide incentives to developers to identify, monitor and address relevant safety- and human rights-related concerns during product design and development and should integrate relevant guidelines into precertification programmes. Regulators should also mandate or conduct robust marketing surveillance to identify biases.

9.6 Policy observatory and model legislation

As AI plays a more prominent role in health systems, governments are introducing national policies and laws to govern its use in health. To ensure that such laws and policies address the ethical concerns and the opportunities associated with use of AI, the OECD launched a policy observatory in 2020 that "aims to help countries enable, nurture and monitor the responsible development of trustworthy artificial intelligence systems for the benefit of society" (*331*).

WHO supports such initiatives and, on the basis of the ethical principles and findings outlined in this report, is exploring collaboration with the OECD on a policy observatory to identify and analyse relevant policies and laws. It is critical that WHO collaborate with other well-placed intergovernmental organizations with wider membership, including of LMIC, such as other United Nations agencies. WHO may also consider issuing model legislation as a reference for governments to develop their own laws to ensure appropriate protection, regulations, rules and safeguards to build the trust of the general public, providers and patients in the use of AI in health-care systems, and, for example, for the management of data and information in ways that improve the accuracy and utility of AI while not compromising privacy, confidentiality or informed consent.

Recommendations

1. WHO should work in a coordinated manner with appropriate intergovernmental organizations to identify and formulate laws, policies and best practices for ethical development, deployment and use of AI technologies for health.

2. WHO should consider issuing model legislation to be used as a reference for governments that wish to build an appropriate legal framework for the use of AI for health.

9.7 Global governance of artificial intelligence

AI is playing an ever-expanding role worldwide. AI has already contributed US$ 2 trillion to global gross domestic product, which could rise to more than US$ 15 trillion by 2030 (*332*). The importance of AI can also be measured by the positive or negative role it might play in achievement of the Sustainable Development Goals. According to one study, AI could enable accomplishment of 134 of the targets but inhibit achievement of 59 targets (*6*).

Ethical principles, regulatory frameworks and national laws on AI continue to proliferate, providing a form of governance; however, the ethical principles and guidance on adherence to international human rights obligations related to AI remain nascent and differ widely among countries, in the public and the private sector and between governments and companies; the platforms of several companies boast more users or subscribers than those of the most populous countries. Thus, company standards influence the control of many AI technologies, including those used in health care.

With the increase in AI standards and laws around the world and diffusion of how and where AI ethics is managed, additional international oversight and enforcement may be necessary to ensure convergence on a core set of principles and requirements that meet ethical principles and human rights obligations. Otherwise, the short-term economic gains that could be made with AI could encourage some governments and companies to ignore ethical requirements and human rights obligations and engage in a "race to the bottom".

First, technical advice from and the engagement of WHO and other intergovernmental organizations such as the Council of Europe, OECD and UNESCO and respect for ethical principles and human rights standards can ensure that companies and governments both move towards common high standards (*333*). In the domain of global health, this will also require that major global health bodies, such as WHO, the Global Fund to Fight AIDS, Tuberculosis and Malaria, United Nations development agencies and foundations, agree on a common position about the risks associated with these technologies and clearly commit themselves to adherence to human rights and ethical standards as a core principle of all strategies and guidance (*333*).

Secondly, global governance could strengthen the voice and role of LMIC, which are less involved in developing AI technologies or in setting international principles. LMIC also lag in use of AI, including in health, partly because of the enduring digital divide, and may not yet have the capacity to regulate use of AI. Thus, global governance could improve access to information and communication and digital technologies in LMIC, guide LMIC governments in accurate assessment of the benefits and risks of AI technologies and hold companies accountable for their practices in LMIC.

Thirdly, global governance could ensure that all governments can adapt to the changes that will be wrought as these technologies become ever more sophisticated and powerful. Independent scientific advice and evidence will be necessary as AI technologies evolve and are translated into policy guidance. For the use of AI for health, it is critical that global health agencies promote only those AI technologies that have been rigorously tested and validated as health interventions by an appropriate authority, such as WHO, and assessed for risks (*333*).

Global governance of use of AI for health will consist partly of adapting governance structures, including the policies and practices of global health agencies, treatment guidelines issued by WHO and global agreements to meet certain health objectives, such as eliminating HIV and AIDS by 2030. Furthermore, global standards should be set for all ethical concerns of AI for health, such as impacts on labour, data governance, privacy, ownership and autonomous decision-making.

As for the use of many other health technologies, nongovernmental organizations and community groups will play critical roles in ensuring that human rights obligations and ethical principles are considered from the onset of decision-making and respected in practice and that governments and companies introduce appropriate safeguards to prevent and respond to any risks and swiftly redress any negative consequences of the use of AI. Civil society and affected communities should participate in the design of AI technologies, and international organizations should work with nongovernmental organizations and affected populations to develop and mainstream guidance for governments and companies.

Several efforts have been made to improve global governance of AI, including the joint initiative of the governments of Canada and France to establish the Global Partnership on AI in June 2020, which now comprises 19 countries. It is intended to convene global AI experts and provide guidance on AI topics, including the future of work, data and privacy (*334*). Its first summit was held in December 2020 (*335*).

Such welcome bilateral and multilateral initiatives should feed into global processes based on the perspectives of all countries. For example, the United Nations Secretary-General's Roadmap for digital cooperation (*336*) recommended in 2019

> creating a strategic and empowered multi-stakeholder high-level body, building on the experience of the existing multi-stakeholder advisory group, which would address urgent issues, coordinate follow-up action on Forum discussions and relay proposed policy approaches and recommendations from the Forum to the appropriate normative and decision-making forums.

Such a multi-stakeholder body would contribute to the wider governance and standard-

setting required for AI and provide means for addressing many of the challenges and questions related to the ethics and governance of the use of AI for health.

Recommendations

1. Governments should support global governance of AI for health to ensure that the development and diffusion of AI technologies is in accordance with the full spectrum of ethical norms, human rights protection and legal obligations.

2. Global health bodies such as WHO, Gavi, the Vaccines Alliance, the Global Fund to Fight AIDS, Tuberculosis and Malaria, Unitaid and major foundations should commit themselves to ensuring that adherence to human rights obligations, legal safeguards and ethical standards is a core obligation of all strategies and guidance.

3. International agencies, such as the Council of Europe, OECD, UNESCO and WHO, should develop a common plan to address the ethical challenges and the opportunities of using AI for health, for example through the United Nations Interagency Committee on Bioethics. The plan should include providing coherent legal and technical support to governments to comply with international ethical guidelines, human rights obligations and the guiding principles established in this report.

4. Governments and international agencies should engage nongovernmental and community organizations, particularly for marginalized groups, to provide diverse insights.

5. Civil society should participate in the design and use of AI technologies for health as early as possible in their conceptualization.

REFERENCES

1. Report of the Secretary-General on SDG progress. Special edition. New York City (NY): United Nations; 2019 (https://sustainabledevelopment.un.org/content/documents/24978Report_of_the_SG_on_SDG_Progress_2019.pdf, accessed 8 November 2020).
2. Timmermans S, Kaufman R. Technologies and health inequities. Ann Rev Sociol. 2020;46:583–602.
3. Report of the Special Rapporteur on the Promotion and protection of the right to freedom and expression. United Nations General Assembly. 73rd Session (A/73/348). New York City (NY): United Nations; 2018 (https://undocs.org/pdf?symbol=en/A/73/348; accessed 7 January 2021).
4. Recommendation of the Council on Artificial Intelligence (OECD Legal Instruments. OECD/LEGAL/O449). Paris: Organization for Economic Co-operation and Development; 2019 (https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449#mainText, accessed 2 December 2020).
5. Hao K. What is machine learning? Machine-learning algorithms find and apply patterns in data. And they pretty much run the world. MIT Technology Revies,17 November 2017 (https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/, accessed 28 August 2020).
6. Vinuesa R, Azizpour H, Leite I, Balaam M, Dignum V, Domisch S et al. The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun. 2020;11:233.
7. Flynn L. When AI is watching patient care: Ethics to consider. Bill of Health, 18 February 2020 (https://blog.petrieflom.law.harvard.edu/2020/02/18/when-ai-is-watching-patient-care-ethics-to-consider/, accessed 13 August 2020).
8. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: How can AI contribute to health in resource-poor settings? BMJ Glob Health. 2018;3:e000798.
9. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet. 2020;395:1579–86.
10. Miller RA, Schaffner KF, Meisel A. Ethical and legal issues related to the use of computer programs in clinical medicine. Ann Intern Med. 1985;102:529–36.
11. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. CA Cancer J Clin. 2019;69(2):127–57.
12. Xiong Y, Ba X, Hou A, Zhang K, Chen L, Li T. Automatic detection of Mycobacterium tuberculosis using artificial intelligence. J Thorac Dis. 2018;10(3):1936–40.
13. Mandavilli A. These algorithms could bring an end to the world's deadliest killer. New York Times. 20 November 2020 (https://nyti.ms/2KnQPu5, accessed 19 January 2021).
14. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. Lancet Digital Health. 2019;1:6.
15. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H et al. Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med. 2018;15(11):1002686.
16. Bejnordi BE, Veta M, van Diest PJ, van Ginneken P, Karssemeijer N, Litjens J et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA. 2017;318(22):2199–210.
17. Alsharqi M, Woodward WJ, Mumith JA, Markham DC, Upton R, Leeson P. Artificial intelligence and echocardiography. Echo Res Pract. 2018;5(4):R115–25.
18. Collis F. Using artificial intelligence to detect cervical cancer. NIH Director's Blog, 17 January 2019 (https://directorsblog.nih.gov/2019/01/17/using-artificial-intelligence-to-detect-cervical-cancer/, accessed 15 February 2021).

19. Innovative, affordable screening and treatment to prevent cervical cancer. Geneva: Unitaid; 2021 (https://unitaid.org/project/innovative-affordable-screening-and-treatment-to-prevent-cervical-cancer/#en, accessed February 2021).
20. Fan R, Zhang N, Yang L, Ke J, Zhao D, Cui Q. AI-based prediction for the risk of coronary heart disease among patients with type 2 diabetes mellitus. Sci Rep. 2020;10:14457.
21. Yan Y, Zhang JW, Zang GY, Pu J. The primary use of artificial intelligence in cardiovascular diseases: What kind of potential role does artificial intelligence play in future medicine? J Geriatr Cardiol. 2019;16(8):585–91.
22. Chaki J, Thillai Ganesh S, Cidham SK, Theertan SA. Machine learning and artificial intelligence based diabetes mellitus detection and self-management: a systematic review. J King Saud Univ Comput Inf Sci. 2020 (https://doi.org/10.1016/j.jksuci.2020.06.013).
23. Singh J. Artificial intelligence and global health: opportunities and challenges. Emerg Topics Life Sci. 2019;3:10.
24. The Topol review: Preparing the healthcare workforce to deliver the digital future. London: National Health Service; 2019 (https://topol.hee.nhs.uk/, accessed 23 August 2020).
25. Hollander JE, Carr BG. Virtually perfect? Telemedicine for COVID-19. N Engl J Med. 2020;382:1679–81.
26. Mou M. COVID-19 gives boost to China's telemedicine industry. Wall Street Journal, 22 October 2020 (https://www.wsj.com/articles/covid-19-gives-boost-to-chinas-telemedicine-industry-11603379296, accessed 3 February 2021).
27. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Health. 2019;5:2055207619871808.
28. Dennis AR, Kim A, Rahimi M, Ayabakan S. User reactions to COVID-19 screening chatbots from reputable providers. J Am Med Informatics Assoc. 2020;27(11):1727–31.
29. Roski J, Chapman W, Heffner J, Trivedi R, Del Fiol G, Kukafka R et al. How artificial intelligence is changing health and health care. In: Matheny M, Thadaney Israni S, Ahmed M, Whicher D, editors. Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington DC: National Academy of Medicine; 2019 (https://nam.edu/artificial-intelligence-special-publication/, accessed 19 July 2020).
30. Marr B. The incredible ways in which artificial intelligence is now used in mental health. Forbes, 3 May 2019 (https://www.forbes.com/sites/bernardmarr/2019/05/03/the-incredible-ways-artificial-intelligence-is-now-used-in-mental-health/?sh=7806594ad02e, accessed 17 May 2020).
31. Gamble A. Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective. Aslib J Inf Manag. 2020;72(4):509–23.
32. What is "biosurveillance"? The COVID-19 measures getting under our skin. Amsterdam: Digital Freedom Fund, 28 May 2020 (https://medium.com/digital-freedom-fund/what-is-biosurveillance-c8bffe70d16f, accessed 17 October 2020).
33. Vincent JL, Moreno R, Takala J, Willatts S, De Mendança A, Bruining H et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Med. 1996;22:707–10.
34. Khanam V, Tusha J, Abkouh DT, Al-Janabi L, Tegeltija V, Kumar S. Sequential organ failure assessment score in patients infected with SARS COV-2. Chest. 2020;158(4):A602.
35. Shickel B, Loftus TJ, Adhikari L, Ozrazgat-Baslanti T, Bihorac A, Rashidi P. DeepSOFA: A continuous acuity score for critically ill patients using clinically interpretable deeplearning. Sci Rep. 2019;9:1879.
36. Shea GP, Solomon CA. Triage in a pandemic: Can AI help ration care? Knowledge@Wharton, 27 March 2020. Philadelphia (PA): University of Pennsylvania (https://knowledge.wharton.upenn.edu/article/triage-in-a-pandemic-can-ai-help-ration-access-to-care/, accessed 3 December 2020)

37. Babic B, Cohen IG, Evgeniou T, Gerke S, Trichakis N. Can AI fairly decide who gets an organ transplant. Harvard Business Review, 1 December 2020. Cambridge (MA): Harvard Business Publishing (https://hbr.org/2020/12/can-ai-fairly-decide-who-gets-an-organ-transplant, accessed 3 December 2020).
38. Raza S. Artificial intelligence for genomic medicine. Cambridge: PHG Foundation, University of Cambridge; 2020 (https://www.phgfoundation.org/documents/artifical-intelligence-for-genomic-medicine.pdf, accessed 11 December 2020).
39. Fleming N. How artificial intelligence is changing drug discovery. Nature Spotlight: Biopharmaceuticals, 30 May 2018 (https://www.nature.com/articles/d41586-018-05267-x, accessed 13 October 2020).
40. New Ebola treatment using artificial intelligence. San Francisco (CA): Atomwise; 2015 (https://www.atomwise.com/2015/03/24/new-ebola-treatment-using-artificial-intelligence/, accessed February 2020).
41. Metz C. London AI lab claims breakthrough that could accelerate drug discovery. The New York Times, 30 November 2020 (https://nyti.ms/2VfKkvA, accessed 14 December 2020).
42. Low LA, Mummery C, Berridge BR, Austin CP, Tagle DA. Organs-on-chips: into the next decade. Nat Rev Drug Discov. 2020 (https://doi.org/10.1038/s41573-020-0079-3, accessed April 2021).
43. Artificial intelligence: How to get it right. London: National Health Service; 2019 (https://www.nhsx.nhs.uk/media/documents/NHSX_AI_report.pdf, accessed 2 August 2020).
44. WHO guidelines on ethical issues in public health surveillance. Geneva: World Health Organization; 2017 (https://apps.who.int/iris/bitstream/handle/10665/255721/9789241512657-eng.pdf;jsessionid=935C008CB079C96FDF2B401C485B49A6?sequence=1, accessed 11 September 2020).
45. Micro-targeting. London: Privacy International; 2021 (https://privacyinternational.org/learn/micro-targeting, accessed 13 January 2021).
46. Smart cities. London: Privacy International; 2021 (https://privacyinternational.org/learn/smart-cities, accessed 15 January 2021).
47. Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457:1012–4.
48. Privacy International and the International Committee for the Red Cross. The humanitarian metadata problem: Doing no harm in the digital era. London: Privacy International and ICRC; 2018 (https://privacyinternational.org/sites/default/files/2018-12/The%20Humanitarian%20Metadata%20Problem%20-%20Doing%20No%20Harm%20in%20the%20Digital%20Era.pdf, accessed 6 February 2021).
49. Cho A. Artificial intelligence systems aim to sniff out signs of COVID-19 outbreaks. Science, 12 May 2020 (https://www.sciencemag.org/news/2020/05/artificial-intelligence-systems-aim-sniff-out-signs-covid-19-outbreaks#, accessed August 2020)
50. Hswen Y, Brownstein JS. Real-time digital surveillance of vaping-induced pulmonary disease. NEJM. 2019;381:1778–80.
51. White RW, Wang S, Pant A, Harpaz R, Shukla P, Sun W et al. Early identification of adverse drug reactions from search log data. J Biomed Informatics. 2016;59:42–8.
52. Precision FDA: Gaining new insights by detecting adverse event anomalies using FDA Open Data. Silver Spring (MD): US Food and Drug Administration; 2020 (https://precision.fda.gov/challenges/9, accessed September 2020).
53. Whitelaw S, Mamas MA, Topol E, Van Spall GC. Applications of digital technology in COVID-19 planning and response. Lancet Digital Health. 2020;2 e435–40.
54. Bullock J, Luccioni A, Pham KH, Nga Lam CS, Luengo-Oroz M. Mapping the landscape of artificial intelligence applications against COVID-19. J Artificial Intell Res. 2020;69:807–45.

55. Toh A. Big Data could undermine the COVID-19 response. Wired, 12 April 2020 (https://www.wired.com/story/big-data-could-undermine-the-covid-19-response/, accessed February 2021).
56. McDonald SM. Ebola: A big data disaster. Privacy, property, and the law of disaster experimentation (CIS Papers 2016.01). Delhi: Centre for Internet and Society; 2016 (https://cis-india.org/papers/ebola-a-big-data-disaster, accessed 20 February 2021).
57. Hao K., Doctors are using AI to triage COVID-19 patients. The tools may be here to stay. MIT Technology Review. 23 April 2020 (https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/, accessed 4 October 2020).
58. AI and control of COVID-19 coronavirus. Strasbourg: Council of Europe; 2020 (https://www.coe.int/en/web/artificial-intelligence/ai-and-control-of-covid-19-coronavirus, accessed 17 September 2020).
59. Ethical considerations to guide the use of proximity tracking technologies for COVID-19 contact tracing. Interim guidance. Geneva: World Health Organization' 2020 (https://www.who.int/publications/i/item/WHO-2019-nCoV-Ethics_Contact_tracing_apps-2020.1, accessed 1 February 2021).
60. Olson, P., Coronavirus reveals limits of AI health tools. Wall Street Journal, 29 February 2020 (https://www.wsj.com/articles/coronavirus-reveals-limits-of-ai-health-tools-11582981201, accessed 5 October 2020).
61. Horowitz BT. Are medical chatbots able to detect coronavirus? Health Tech Magazine, 10 September 2020 (https://healthtechmagazine.net/article/2020/09/are-medical-chatbots-able-detect-coronavirus, accessed 28 October 2020).
62. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1:389–99.
63. Question of the realization of economic, social and cultural rights in all countries: the role of new technologies for the realization of economic, social and cultural rights: Report of the Secretary General. Geneva: Office of the High Commissioner for Human Rights; 2020 (https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session43/Documents/A_HRC_43_29.pdf, accessed 9 January 2021)
64. Secretary-General Guterres calls for a global reset to recover better, guided by human rights. Geneva: United Nations Human Rights Council; 2021 (https://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=26769&LangID=E, accessed 3 March 2021).
65. The Toronto Declaration. Protecting the right to equality and non-discrimination in machine learning systems. Amnesty International and Access Now; 2018 (https://www.torontodeclaration.org/declaration-text/english/, accessed 4 June 2020).
66. Addressing the impact of algorithms on human rights. Strasbourg: Council of Europe' 2019 (https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf, accessed 16 December 2020).
67. European Convention on Human Rights. Strasbourg: Council of Europe; 2010 (https://www.echr.coe.int/documents/convention_eng.pdf, accessed 6 March 2021)
68. Convention for the Protection of Human Rights and Dignity of the Human Being with Regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. Strasbourg: Council of Europe; 1997 (https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168007cf98, accessed 3 March 2020).
69. Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data. Strasbourg: Council of Europe; 1981 (https://rm.coe.int/1680078b37, accessed 13 April 2020).
70. Guidelines on artificial intelligence and data protection. Strasbourg: Council of Europe; 2019 (https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8, accessed 13 April 2020).
71. European ethical charter on the use of artificial intelligence in judicial systems and their environment. Strasbourg: Council of Europe; 2018 (https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c, accessed 20 April 2020).

72. General recommendations for the processing of personal data in artificial intelligence. Brussels: Red IberoAmerica de Proteccion de Datos, European Union; 2019 (https://www.redipd.org/sites/default/files/2020-02/guide-general-recommendations-processing-personal-data-ai.pdf, accessed 27 October 2020).
73. Specific guidelines for compliance with the principles and rights that govern the protection of personal data in artificial intelligence projects. Brussels: Red IberoAmerica de Proteccion de Datos, European Union; 2019 (https://www.redipd.org/sites/default/files/2020-02/guide-specific-guidelines-ai-projects.pdf, accessed 27 October 2020).
74. Recommendation CM/Rec (2019)2 of the Committee of Ministers to Member States on the protection of health-related data. Strasbourg: Council of Europe; 2019 (https://www.apda.ad/sites/default/files/2019-03/CM_Rec%282019%292E_EN.pdf, accessed 14 April 2020).
75. African Union Convention on Cyber Security and Personal Data Protection. Addis Ababa: African Union; 2014 (https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection, accessed 19 February 2021).
76. Internet Society, Commission of the African Union. Personal data protection guidelines for Africa. Reston (VA): Internet Society; 2018 (https://www.internetsociety.org/resources/doc/2018/personal-data-protection-guidelines-for-africa/, accessed 19 February 2021)
77. The digital transformation strategy for Africa (2020–2030). Addis Ababa: African Union; 2020 (https://au.int/sites/default/files/documents/38507-doc-dts-english.pdf, accessed 12 February 2021).
78. Recommendations for data and biospecimen governance in Africa. Nairobi: African Academy of Sciences; 2021 (https://www.aasciences.africa/sites/default/files/Publications/Recommendations%20for%20Data%20and%20Biospecimen%20Governance%20in%20Africa.pdf, accessed 26 February 2021).
79. Zeng Y, Lu E, Huangfu C. Linking artificial intelligence principles. In: Proceedings of the AAAI Workshop on Artificial Intelligence Safety, Honolulu, Hawaii, 2019. Aachen: CEUR Workshop Proceedings; 2019 (https://arxiv.org/ftp/arxiv/papers/1812/1812.04814.pdf, accessed 12 February 2020).
80. OECD legal instruments. Recommendations of the Council on artificial Intelligence. Paris: Organization for Economic Co-operation and Development; 2019 (https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL0449, accessed 12 February 2020).
81. Going digital. Making the transformation work for growth and well-being. Paris: Organization for Economic Co-operation and Development; 2019 (https://www.oecd.org/going-digital/ai/principles/, accessed 5 February 2020).
82. Economy, employment, and education in the digital age. What can the G20 do to implement AI principles and to shape global data governance? Berlin: Global Solutions Initiative; 2021 (https://www.global-solutions-initiative.org/global-table/ai-and-data-governance/#:~:text=Under%20Japan%E2%80%99s%20presidency%2C%20the%20G20%20endorsed%20Principles%20for,pursuit%20of%20beneficial%20outcomes%20for%20people%20and%20planet.%E2%80%9D, accessed April 2021).
83. OECD AI policy observatory. Paris: Organization for Economic Co-operation and Development; 2019 (https://www.oecd.org/going-digital/ai/about-the-oecd-ai-policy-observatory.pdf, accessed 5 February 2020).
84. Unboxing artificial intelligence: 10 steps to protect human rights. Strasbourg: Council of Europe; 2019 (https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64, accessed 6 February 2020).
85. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019 (https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai, accessed 13 February 2020).

86. AI utilisation guidelines. The Conference Towards AI Society. Paris: Organization for Economic Co-operation and Development; 2019 (https://www.oecd.ai/dashboards/policy-initiatives/2019-data-policyInitiatives-24346, accessed 5 February 2021).
87. Governance principles for the new generation artificial intelligence – Developing responsible artificial intelligence. China Daily, 17 June 2019 (http://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html, accessed 20 April 2021).
88. Beijing AI principles. Beijing: Beijing Academy of Artificial Intelligence; 25 May 2019. (https://www.baai.ac.cn/news/beijing-ai-principles-en.html, accessed 20 April 2021)
89. Singapore wins international award for its artificial intelligence governance and ethics initiatives. Singapore: InfoComm Media Development Authority, 9 April 2019 (https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2019/singapore-wins-international-award-for-its-artificial-intelligence-governance-and-ethics-initiatives, accessed 16 February 2020).
90. Singapore Computer Society, InfoComm Media Development Authority. AI Ethics & Governance Body of Knowledge. Singapore: Singapore Computer Society; 2020 (https://ai-ethics-bok.scs.org.sg/about, accessed 9 December 2020).
91. African Union High Level Panel on Emerging Technologies (APET). Addis Ababa: African Union Development Agency; 2019 (https://www.nepad.org/microsite/african-union-high-level-panel-emerging-technologies-apet, accessed 17 January 2021).
92. Declaration of Astana. Global Conference on Primary Health Care, Astana, 25–26 October 2018. Geneva: World Health Organization; 2018 (https://www.who.int/docs/default-source/primary-health/declaration/gcphc-declaration.pdf, accessed 14 February 2020).
93. International Bioethics Committee. Report of the IBC on big data and health. Paris: United Nations Educational, Cultural and Scientific Organization; 2017 (https://unesdoc.unesco.org/ark:/48223/pf0000248724, accessed 20 February 2020).
94. World Commission on the Ethics of Scientific Knowledge and Technology. Report of COMEST on robotics ethics. Paris: United Nations Educational, Cultural and Scientific Organization; 2017 (https://unesdoc.unesco.org/ark:/48223/pf0000253952, accessed 20 February 2020).
95. Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of artificial intelligence. Paris: United Nations Educational, Cultural and Scientific Organization; 2019 (https://unesdoc.unesco.org/ark:/48223/pf0000367422, accessed 20 February 2020).
96. Guidance. Code of conduct for data-driven health and care technology. London: Department of Health and Social Care; 2019 (https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology, accessed 20 February 2020).
97. The Lancet, Financial Times Commission. Governing health futures 2030: Growing up in a digital world. Geneva: Global Health Centre, The Graduate Institute; 2021 (https://www.governinghealthfutures2030.org/#:~:text=For%20the%20first%20time%2C%20a,support%20attainment%20of%20the%20third, accessed 4 March 2021).
98. French bioethics law: an original participatory approach for the National Bioethics Consultation. Paris: Institut Pasteur, 2 September 2019 (https://www.pasteur.fr/en/home/research-journal/reports/french-bioethics-law-original-participatory-approach-national-bioethics-consultation, accessed 16 April 2021).
99. Ross WD. The right and the good. Oxford: Clarendon Press; 1930.
100. Beauchamp TL, Childress JF. The principles of biomedical ethics. 5th edition. New York City (NY): Oxford University Press; 2001.

101. Public debate. Strasbourg: Council of Europe; 2021 (https://www.coe.int/en/web/bioethics/public-debate, accessed 17 August 2020).
102. Morozov E. To save everything, click here. New York City (NY): Public Affairs; 2014.
103. Matheny M, Thadaney Israni S, Ahmed M, Whicher D, editors. Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington DC: National Academy of Medicine; 2019 (https://nam.edu/artificial-intelligence-special-publication/, accessed 18 November 2020).
104. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: Taxonomy, ethical challenges, and navigation aid. Lancet Digit Health. 2020;2(8):e425–34.
105. Fenech M, Strukelj N, Buston O. The ethical, social, and political challenges of artificial intelligence in healthcare. London: Future Advocacy; 2018 (https://cms.wellcome.org/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf, accessed November 2020;
106. London AJ. Groundhog day for medical artificial intelligence. Hastings Centre Rep. 2018;;48(3): doi: 10.1002/hast.842.
107. In tech-driven 21st century, achieving global development goals requires closing digital gender divide. UN News, 15 March 2019 (https://news.un.org/en/story/2019/03/1034831, accessed 16 November 2020).
108. The age of digital interdependence: Report of the United Nations Secretary-General's High-level Panel on Digital Cooperation. New York City (NY): United Nations; 2019 (https://www.un.org/en/pdfs/HLP%20on%20Digital%20Cooperation%20Report%20Executive%20Summary%20-%20ENG.pdf, accessed 14 November 2020).
109. Schwerhoff G, Sy M. Where the sun shines. Washington DC: International Monetary Fund Finance and Development; 2020 (https://www.imf.org/external/pubs/ft/fandd/2020/03/pdf/powering-Africa-with-solar-energy-sy.pdf, accessed 11 February 2021).
110. SDG7: Access to affordable, reliable, sustainable and modern energy for all. Paris: International Energy Agency; 2020 (https://www.iea.org/reports/sdg7-data-and-projections/access-to-electricity, accessed 29 November 2020).
111. Winslow J. America's digital divide. Pew Trust Magazine, 26 July 2019 (https://www.pewtrusts.org/en/trust/archive/summer-2019/americas-digital-divide, accessed 12 September 2020).
112. Buying a smartphone on the cheap? Privacy might have to be the price you have to pay. London: Privacy International; 2019 (https://privacyinternational.org/long-read/3226/buying-smart-phone-cheap-privacy-might-be-price-you-have-pay, accessed 23 February 2021).
113. Vayena E, Blassime A. Biomedical big data: New models of control over access, use, and governance. Bioethical Inquiry. 2017;14:501–13.
114. Evolving health data ecosystem. Geneva: World Health Organization; 2016 (https://www.who.int/ehealth/resources/ecosystem.pdf?ua=1, accessed 1 March 2021).
115. Vayena E, Dzenowagis J, Langfeld M. Evolving health data ecosystem. Geneva: World Health Organization; 2016 (https://www.who.int/ehealth/resources/ecosystem.pdf?ua=1, accessed 17 April 2021).
116. McNair D, Price WN. Health care AI: Law, regulation, and policy. In: Matheny M, Thadaney Israni S, Ahmed M, Whicher D, editors. Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington DC: National Academy of Medicine; 2019.
117. Xafis V, Schaefer GO, Labude MK, Brassington I, Ballantyne A, Lim HY et al. An ethics framework for big data in health and research. Asian Bioethics Rev. 2019;11:227–54.
118. White paper: On artificial intelligence – A European approach to excellence and trust. Brussels: European Commission; 2020 (https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf, accessed 8 November 2020).

119. Mozur P, Zhong R, Krolik A. In coronavirus fight, China gives citizens a color code, with red flags. The New York Times, 1 March 2020 (https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html, accessed 14 June 2020).
120. Angwadi centres, digital tracking in India's blueprint for COVID-19 vaccine drive. The Wire (Science), 6 November 2020 (https://science.thewire.in/health/anganwadi-centres-digital-tracking-in-indias-blueprint-for-covid-19-vaccination-drive/, accessed 13 January 2021).
121. Immunity passports and COVID-19: An explainer. London: Privacy International; 2020 (https://privacyinternational.org/explainer/4075/immunity-passports-and-covid-19-explainer, accessed 30 November 2020).
122. Fisher M, Han CS. How South Korea flattened the curve. The New York Times, 23 March 2020 (https://www.nytimes.com/2020/03/23/world/asia/coronavirus-south-korea-flatten-curve.html, accessed 7 December 2020).
123. The looming disaster of immunity passports and digital identity. London: Privacy International; 2020 (https://privacyinternational.org/long-read/4074/looming-disaster-immunity-passports-and-digital-identity, accessed 12 November 2020).
124. A fair shot: Ensuring universal access to COVID-19 diagnostics, treatments, and vaccines. London: Amnesty International; 2020 (https://www.amnesty.org/download/Documents/POL3034092020ENGLISH.PDF, accessed 16 December 2020).
125. Zuboff S. The age of surveillance capitalism. London: Principle Books; 2019.
126. Illmer, Andreas, Singapore reveals COVID privacy data available to police. BBC News, 5 January 2021 (https://www.bbc.com/news/world-asia-55541001, accessed 11 February 2021).
127. Chee K. Bill introduced to make clear TraceTogether, SafeEntry data can be used to look into only 7 types of serious crimes. Straits Times, 1 February 2021 (https://www.straitstimes.com/singapore/proposed-restrictions-to-safeguard-personal-contact-tracing-data-will-override-all-other, accessed 22 February 2021).
128. Price WN II, Cohen IG. Privacy in the age of medical big data. Nature Med. 2019;25(1):37–43.
129. Copeland R. Google's Project Nightingale gathers personal health data on millions of Americans. Wall Street Journal, 11 November 2019 (https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790, accessed 12 November 2020).
130. Wood M. U Chicago Medicine collaborates with Google to use machine learning for better health care. At the Forefront: U Chicago Medicine, 17 May 2017 (https://www.uchicagomedicine.org/forefront/research-and-discoveries-articles/uchicago-medicine-collaborates-with-google-to-use-machine-learning-for-better-health-care, accessed 19 January 2021).
131. Shachar C, Gerke S, Minssen T. Is data sharing caring enough about patient privacy? Part I: The background. Cambridge (MA): Bill of Health, Harvard Law, Petrie Flom Center; 2019 (https://blog.petrieflom.law.harvard.edu/2019/07/26/is-data-sharing-caring-enough-about-patient-privacy-part-i-the-background/, accessed 13 March 2021).
132. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M et al. Scalable and accurate deep learning with electronic health records. npj Digital Med. 2018;1:18.
133. Andanda P. Ethical and legal governance of health-related research that use digital data from user-generated online health content. Inf Commun Soc. 2020;23(8):1154–69.
134. Fussell S. Google's totally creepy, totally legal health-data harvesting. The Atlantic, 14 November 2019 (https://www.theatlantic.com/technology/archive/2019/11/google-project-nightingale-all-your-health-data/601999/, accessed 30 November 2020).

135. Lewis P, Conn D, Pegg D. UK government using confidential patient data in coronavirus response. The Guardian, 12 April 2020 (https://www.theguardian.com/world/2020/apr/12/uk-government-using-confidential-patient-data-in-coronavirus-response, accessed 30 November 2020).
136. Hern A. Anonymous browsing data can be easily exposed, researchers reveal. The Guardian, 1 August 2017 (https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers, accessed 12 February 2020).
137. Ross C. At Mayo Clinic, sharing patient data with companies fuels AI innovation – and concerns about consent. STAT News, 3 June 2020 (https://www.statnews.com/2020/06/03/mayo-clinic-patient-data-fuels-artificial-intelligence-consent-concerns/, accessed 18 November 2020).
138. Mann L. Left to other peoples' devices? A political economy perspective on the Big Data revolution in development. Dev Change. 2017;49(2):doi: 10.1111/dech.12347.
139. Hariri Y. How to survive the 21st century. Geneva: World Economic Forum; 2020 (https://www.weforum.org/agenda/2020/01/yuval-hararis-warning-davos-speech-future-predications/, accessed 18 November 2020).
140. Krutzinna J, Taddeo M, Floridi L. Enabling posthumous medical data donation: A plea for the ethical utilisation of personal health data. In: Krutzinna J, Floridi L, editors, The ethics of medical data donation (Philosophical Studies Series, Vol 137). Cham: Springer; 2019
141. Shaw DM, Gross JV, Erren TC. Why you should donate your health data (as well as your organs) when you die. STAT News, 14 February 2017. (https://www.statnews.com/2017/02/14/donate-health-data-death/, accessed 18 September 2020).
142. General Data Protection Regulation. Article 27. Brussels: European Union; 2016 (https://eur-lex.europa.eu/eli/reg/2016/679/oj, accessed 18 March 2021).
143. Malgieri G. RIP: Rest in privacy or rest in (quasi-)property? Personal data protection of deceased data subjects between theoretical scenarios and national solutions. In: Leenes R, van Brackel R, Gutwirth S, De Hert P, editors. Data protection and privacy: The Internet of bodies. Brussels: Hart; 2018) (https://ssrn.com/abstract=3185249, accessed 12 February 2021).
144. At a glance: De-identification, anonymisation, and pseudo-anonymisation under the GDPR. Boulder (CO): Bryan Cave Leighton Paisner; 2017 (https://www.bclplaw.com/en-US/insights/at-a-glance-de-identification-anonymization-and-pseudonymization-1.html, accessed 12 September 2020)
145. General Data Protection Regulation, Article 5. Brussels: European Union; 2016 (https://eur-lex.europa.eu/eli/reg/2016/679/oj, accessed 18 March 2021).
146. Bari L, O'Neill D. Rethinking patient data privacy in the era of digital health. Health Affairs, 12 December 2019 (https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/, accessed 23 November 2020).
147. Rocher L, Hendrickx JM, de Montjoye Y. Estimating the success of re-identifications in incomplete datasets using generative models. Nat Commun. 2019;10:3069.
148. May T. Sociogenetic risks – ancestry DNA testing, third-party identity, and protection of privacy. NEJM. 2018;379:410–2.
149. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics. 2020;46(3):205–11.
150. Yeung K. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework. Strasbourg: Council of Europe; 2019 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3286027, accessed 15 March 2020).
151. Habli I, Lawton T, Porter Z. Artificial intelligence in healthcare: Accountability and safety. Bull World Health Organ. 2020;98:251–6.

152. Hurtgen H, Kerkhoff S, Lubatschowski J, Möller M. Rethinking AI talent strategy as automated machine learning comes of age. New York City (NY): McKinsey and Co., 2020 (https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/rethinking-ai-talent-strategy-as-automated-machine-learning-comes-of-age, accessed 4 November 2020).
153. Dixon-Woods M, Pronovost PJ. Patient safety and the problem of many hands. BMJ Qual Saf. 2016;25(7):485–8.
154. Van de Poel I, Royakkers L, Zwart SD, de Lima T, Doorn N, Fahlquist JN. Moral responsibility and the problem of many hands. New York City (NY): Routledge; 2015.
155. Braun M, Hummel P, Beck S, Dabrock P. Primer on an ethics of AI-based decision support systems in the clinic. J Med Ethics. 2020;doi:10.1136/medethics-2019-105860.
156. Metcalf J, Moss E, Boyd D. Owning ethics: Corporate logics, Silicon Valley, and the institutionalisation of ethics. Soc Res. 2019;82(2):449–76.
157. Whitaker M, Crawford K, Dobbe R, Fried G, Kaziunas E, Mathur V et al. AI Now report 2018. New York City (NY): AI Now Institute; 2018 (https://ainowinstitute.org/AI_Now_2018_Report.pdf, accessed 13 December 2020).
158. Vincent J. The problem with ethics: Is Big Tech's embrace of AI ethics boards actually helping anyone. The Verge, 3 April 2019 (https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech, accessed 14 January 2021).
159. Artificial intelligence in healthcare. London: Academy of Medical Royal Colleges; 2019 (https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/, accessed 18 July 2020).
160. In brief: Artificial intelligence in health care. Stockholm: Swedish National Council on Medical Ethics; 2020 (https://smer.se/en/publications/?date=2020-5, accessed 12 July 2020).
161. Duran JM. Computer simulations in science and engineering. Cham: Springer; 2018 (https://link.springer.com/book/10.1007%2F978-3-319-90882-3, accessed April 2021).
162. Humphreys P. The philosophical novelty of computer simulation methods. Synthese. 2009;169(3):615–26.
163. Topol E. Twitter, 7 January 2019 (https://twitter.com/EricTopol/status/1082363519675248640/photo/1, accessed 15 March 2021).
164. Pasquale F. The black box society: The secret algorithms that control money and Information. Camridge (MA): Harvard University Press; 2015.
165. London AJ. Artificial intelligence and black-box medical decisions: Accuracy versus explainability. Hastings Center Rep. 2019;49(1):15–21.
166. Durán JM, Formanek N. Grounds for trust: Essential epistemic opacity and computational reliabilism. Minds Machines. 2018;28:645–66.
167. Cohen IG. Informed consent and medical artificial intelligence: What to tell the patient? Georgetown Law J. 2020;108:1425.
168. Minssen T, Rajam N, Bogers M. Clinical trial data transparency and GDPR compliance: Implications for data sharing and open innovation. Sci Public Policy. scaa014: doi.org/10.1093/scipol/scaa014.
169. Mullin E. Healthcare is the next battleground for Big Tech. One Zero, 27 January 2020 (https://onezero.medium.com/health-care-is-the-next-battleground-for-big-tech-477a7263974, accessed 14 January 2021).
170. Turea M. How the big 4 tech companies are leading innovation. Healthcare Weekly, 27 February 2019 (https://healthcareweekly.com/how-the-big-4-tech-companies-are-leading-healthcare-innovation/, accessed 15 January 2021).
171. A look back at Alphabet's moves in 2019. MobiHealthNews, 13 December 2019 (https://healthcareweekly.com/how-the-big-4-tech-companies-are-leading-healthcare-innovation/, accessed 12 November 2020).

172. Our work with Google Health UK. London: NHS Royal Free London; 2019 (https://www.royalfree.nhs.uk/patients-visitors/how-we-use-patient-information/our-work-with-deepmind/, accessed 12 February 2021).
173. Shepherd C. China's online health platforms boom in wake of coronavirus. The Financial Times, 16 December 2020 (https://www.ft.com/content/22b22543-0fb5-4a8a-8ec0-e3fd067a5190, accessed 1 February 2021).
174. Bridging gaps in healthcare industry with technology. Shenzhen: Tencent Holdings Ltd; 2019 (https://www.tencent.com/en-us/articles/2200933.html, accessed 6 November 2020).
175. How Baidu, Alibaba, and Tencent aim to disrupt Chinese health care. Forkast, 28 January 2020 (https://forkast.news/baidu-alibaba-tencent-china-health-care-blo/, accessed 13 February 2021).
176. Jourdan A. AI ambulances and robot doctors: China seeks digital salve to ease hospital strain. Reuters, 28 June 2018 (https://de.reuters.com/article/us-china-healthcare-tech-idUKKBN1JO1VB, accessed 24 November 2020).
177. Goodman K. Ethics, medicines, and information technology: Intelligent machines and the transformation of health care. Cambridge: Cambridge University Press; 2016.
178. Chen C. Only seven of Stanford's first 5000 vaccines were designated for medical residents. ProPublica, 18 December 2020 (https://www.propublica.org/article/only-seven-of-stanfords-first-5-000-vaccines-were-designated-for-medical-residents, accessed 18 February 2021).
179. Hariri YN. Homo deus: A brief history of tomorrow. London: Vintage; 2015.
180. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW et al. A deep learning model to predict a diagnosis of Alzheimer disease by using 18F-FDG PET of the brain. Radiology. 2019;290(2):456–64.
181. Chen JH, Beam A, Saria S, Mendonça E. Potential trade-offs and unintended consequences of AI. In: Matheny M, Thadaney Israni S, Ahmed M, Whicher D, editors. Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington DC: National Academy of Medicine; 2019.
182. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A et al. A clinically applicable approach to continuous prediction of future acute kidney injury. Nature. 2019;572:116–9.
183. Morley J, Caio C, Machado V, Burr C, Cowls J, Joshi I et al. The debate on the ethics of AI in health care: a reconstruction and critical review. Oxford: Oxford Internet Institute; 2019 (https://digitalethicslab.oii.ox.ac.uk/wp-content/uploads/sites/87/2019/11/The-Debate-on-the-EThics-of-AI-in-Health-Care-pre-print-.pdf, accessed 15 July 2020).
184. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015;26(1):15030.
185. Marcus J, Hurley L, Krakower D, Alexeeff S, Silverberg M, Volk J. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: a modelling study. Lancet HIV. 2019;6:10.1016/S2352-3018(19)30137-7.
186. Urtubey y una insólita propuesta de "prever" embarazos adolescents [Urtubey and an unusual proposal to "anticipate" adolescent pregnancies]. Diario de Cuyo, 11 April 2018 (https://www.diariodecuyo.com.ar/argentina/Urtubey-y-una-insolita-propuesta-de-prever-embarazos-adolescentes-20180411-0081.html, accessed 12 February 2021).
187. Peña P, Varon J. Decolonising AI: A transfeminist approach to data and social justice. In: Artificial intelligence: Human rights, social justice and development. Global Information Society Watch. Association for Progressive Communications; 2019 (https://giswatch.org/sites/default/files/gisw2019_web_th4.pdf, accessed 12 February 2021).

188. Sobre la predicción automática de embarazos adolescents [On automatic prediction of adolescent pregnancies]. Buenos Aires: Universidad de Buenos Aires, Laboratorio de Inteligencia Artificial Aplicada; 2018 (https://www.dropbox.com/s/r7w4hln3p5xum3v/%5BLIAA%5D%20Sobre%20la%20predicci%C3%B3n%20autom%C3%A1tica%20de%20embarazos%20adolescentes.pdf, accessed 12 February 2021).
189. Venturini J. Surveillance and social control: How technology reinforces structural inequality in Latin America. London: Privacy International; 2019 (https://privacyinternational.org/news-analysis/3263/surveillance-and-social-control-how-technology-reinforces-structural-inequality, accessed 12 February 2021).
190. Ortiz Freuler J, Iglesias C. Algorithms and artificial intelligence in Latin America: A study of implementation by governments in Argentina and Uruguay. Washington DC: World Wide Web Foundation; 2018 (http://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Screen_AW.pdf, accessed 12 February 2021).
191. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.
192. Price WN II. Medical AI and contextual bias (U of Michigan Public Law Research Paper No. 632). Harvard J. Law Technol. 2019;66 (https://ssrn.com/abstract=3347890).
193. Minssen T, Gerke S, Aboy M, Price N, Cohen G. Regulatory responses to medical machine learning J Law Biosci. lsaa002 (https://doi.org/10.1093/jlb/lsaa002).
194. Gerke S, Minssen T, Yu H, Cohen IG. Ethical and legal issues of ingestible electronic sensors. Nat Electron. 2019;2:329–34.
195. Simonite T. How an algorithm blocked kidney transplants to patients. Wired Magazine, 26 October 2020 (https://www.wired.com/story/how-algorithm-blocked-kidney-transplants-black-patients/, accessed 12 November 2020).
196. Benjamin R. Assessing risk, automating racism. Science. 2019;366(6464):421–2.
197. Lashbrook A. AI-driven dermatology could leave dark-skinned patients behind. The Atlantic, 16 August 2018 (https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/, accessed 14 December 2020).
198. Bridging the digital gender divide: Include, upskill, innovate. Paris: Organization for Economic Co-operation and Development; 2018 (http://www.oecd.org/digital/bridging-the-digital-gender-divide.pdf, accessed 3 November 2020).
199. Munshi N. How unlocking the secrets of African DNA could change the world. The Financial Times, 5 March 2020 (https://www.ft.com/content/eed0555c-5e2b-11ea-b0ab-339c2307bcd4, accessed 2 November 2020).
200. Devlin H., Genetics research "biased towards studying white Europeans". The Guardian, 8 October 2018 (https://www.theguardian.com/science/2018/oct/08/genetics-research-biased-towards-studying-white-europeans, accessed 2 November 2020).
201. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. npj Digital Med. 2020;3:81.
202. Rose E. How self-tracking apps exclude women. The Atlantic, 15 December 2014 (https://www.theatlantic.com/technology/archive/2014/12/how-self-tracking-apps-exclude-women/383673/, accessed 30 January 2021).
203. Hern A. Fault in NHS Covid app meant thousands at risk did not quarantine. The Guardian, 2 November 2020 (https://www.theguardian.com/world/2020/nov/02/fault-in-nhs-covid-app-meant-thousands-at-risk-did-not-quarantine, 20 January 2020).

204. Contag M, Li G, Pawlowski A, Domke F, Levchenko K, Holz T et al. How they did it: An analysis of emission defeat devices in modern automobiles. In: IEEE Symposium on Security and Privacy. San Diego (CA): University of California at San Diego; 2017 (https://cseweb.ucsd.edu/~klevchen/diesel-sp17.pdf, accessed 8 November 2020).
205. Baraniuk C. How tech bugs could be killing thousands in our hospitals. New Scientist, 16 May 2018 (https://www.newscientist.com/article/mg23831781-700-how-tech-bugs-could-be-killing-thousands-in-our-hospitals/, accessed 20 August 2020).
206. Xu K, Soucat A, Kutzin J, Siroka A, Aranguren Garcia M, Dupuy J et al. Global spending on health: A world in transition. Geneva; World Health Organization; 2019 (https://apps.who.int/iris/bitstream/handle/10665/330357/WHO-HIS-HGF-HF-WorkingPaper-19.4-eng.pdf?ua=1, accessed 25 February 2021).
207. Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. Swiss Med Wkly. 2018;148:w14571
208. de Montjoye YA, Hidalgo CA, Verleysen M, Blondel VD. Unique in the crowd: The privacy bounds of human mobility. Sci Rep. 2013;3:1376.
209. Telemedicine: Opportunities and developments in Member States (Global Observatory for eHealth Series, Vol. 2). Geneva: World Health Organization; 2010 (https://www.who.int/goe/publications/goe_telemedicine_2010.pdf, accessed 17 February 2021).
210. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Health J. 2019;6(2):94–8.
211. Health workforce. Geneva: World Health Organization; 2021 (https://www.who.int/health-topics/health-workforce#tab=tab_1, accessed 7 December 2020).
212. Parikh RK. Should doctors play along with the uberization of health care. Slate, 14 June 2017 (https://slate.com/technology/2017/06/should-doctors-play-along-with-the-uberization-of-health-care.html, accessed 11 November 2020).
213. Gawande A. Why doctors hate their computers. The New Yorker, 5 November 2018 (https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers, accessed 12 December 2020).
214. COVID-19 response: Corporate exploitation. London: Privacy International; 2020 (https://privacyinternational.org/news-analysis/3592/covid-19-response-corporate-exploitation, accessed 13 February 2021).
215. Powles J, Hodson H. Google Deepmind and healthcare in an age of algorithms. Health Technol (Berl). 2017;7(4):351–67.
216. Ballantyne A, Stewart C. Big data and public–private partnerships on healthcare and research. Asian Bioethics Rev. 2019;11:315–26.
217. Hodson H. Revealed: Google AI has access to huge haul of NHS patient data. New Scientist, 29 April 2016 (https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/, accessed 29 November 2020).
218. Durkee A. Facebook to pay millions for allegedly mishandling user data (again). Vanity Fair, 30 January 2020 (https://www.vanityfair.com/news/2020/01/facebook-settlement-facial-recognition-illinois-privacy, accessed 29 November 2020).
219. Mergers: Commission clears acquisition of Fitbit by Google, subject to conditions. Brussels: European Commission; 2020. (https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2484, accessed 12 February 2021).
220. Mergers: Commission opens in-depth investigation into the proposed acquisition of Fitbit by Google. Brussels: European Commission; 2020 (https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1446, accessed 12 November 2020).
221. Competition and data. London: Privacy International; 2021 (https://privacyinternational.org/learn/competition-and-data, accessed 6 February 2021).

222. Bridging gaps in healthcare industry with technology. Shenzhen: Tencent Holdings Ltd; 2019 (https://www.tencent.com/en-us/articles/2200933.html, accessed 24 November 2020).
223. Veale M. Privacy is not the problem with the Google-Apple contact-tracing toolkit. The Guardian, 1 July 2020 (https://www.theguardian.com/commentisfree/2020/jul/01/apple-google-contact-tracing-app-tech-giant-digital-rights, accessed 24 November 2020).
224. Hao K. Training an AI model can emit as much carbon as five cars in their lifetime. MIT Technology Review, 6 June 2019 (https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/, accessed 12 December 2020)
225. DeWeerdt S. It's time to talk about the carbon footprint of artificial intelligence. Anthropocene, 10 November 2020 (https://www.anthropocenemagazine.org/2020/11/time-to-talk-about-carbon-footprint-artificial-intelligence/, accessed 27 February 2021).
226. Climate change. Geneva: World Health Organization; 2021 (https://www.who.int/health-topics/climate-change#tab=tab_1, accessed 27 February 2021).
227. Hao K. We read the paper that forced Timnit Gebru out of Google. Here's what it says. MIT Technology Review, 4 December 2020 (https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/?utm_source=Nature+Briefing&utm_campaign=ebee85e120-briefing-dy-20201208&utm_medium=email&utm_term=0_c9dfd39373-ebee85e120-44944633, accessed 12 December 2020).
228. van den Hoven J, Vermaas PE, van de Poel I, editors. Handbook of ethics, values, and technological design: Sources, theories, values, and application domains. Cham: Springer; 2015 (https://www.springer.com/gp/book/9789400769694#aboutAuthors, accessed April 2021).
229. Aizenberg E, van den Hoven J. Designing for human rights in AI. Big Data Society. 2020;July–December:1–14.
230. Statement regarding the ethical implementation of artificial intelligence systems (AIS) for addressing the COVID-19 pandemic. New York City (NY): Institute of Electrical and Electronic Engineering; 2020 (https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/gieais-covid.pdf?utm_medium=undefined&utm_source=undefined&utm_campaign=undefined&utm_content=undefined&utm_term=undefined, accessed 1 November 2020).
231. Independent High-level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019 (https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top, accessed 12 November 2020).
232. Whitaker K. Citizen science platform with Autistica. London: The Alan Turing Institute; 2019 (https://www.turing.ac.uk/research/research-projects/citizen-science-platform-autistica, accessed 12 November 2020).
233. The elements of informed consent: A toolkit. V.3. Seattle (WA): Sage Bionetworks; 2020 (https://sagebionetworks.org/wp-content/uploads/2020/01/SageBio_EIC-Toolkit_V3_21Jan20_final.pdf, accessed 13 November 2020).
234. EPI-Brain webpage. Geneva: World Health Organization; 2020 (https://www.epi-brain.com/, accessed 3 November 2020).
235. IEEE P7000. IEEE draft model process for addressing ethical concerns during system design. Piscataway (NJ): IEEE Standards Association; 2016 (standards.ieee.org/project/7000.html, accessed April 2021).
236. Understanding patient data. London: Wellcome Trust; 2019 (https://understandingpatientdata.org.uk/, accessed 13 February 2020).
237. Sharing anonymised patient-level data where there is a mixed public and private benefit – a new report. London: Health Research Authority; 2019 (https://www.hra.nhs.uk/about-us/news-updates/sharing-anonymised-patient-level-data-where-there-mixed-public-and-private-benefit-new-report/, accessed 17 February 2020).

238. Artificial intelligence and health, summary report of a roundtable held on 16 January 2019. London: Academy of Medical Sciences; 2019 (https://acmedsci.ac.uk/file-download/77652269, accessed 17 February 2020).
239. Our data driven future in healthcare. People and partnerships at the heart of health-related technologies. London: Academy of Medical Sciences; 2018 (https://acmedsci.ac.uk/file-download/74634438, accessed 17 February 2020).
240. Trust in technology. London: HSBC Holdings Ltd; undated (http://www.hsbc.com/trust-in-technology-report, accessed April 2021).
241. Trustworthy AI in health: Background paper for the G20 AI dialogue, digital economy, and trade. Paris: Organization for Economic Co-operation and development; 2020 (https://www.oecd.org/health/trustworthy-artificial-intelligence-in-health.pdf, accessed 11 November 2020).
242. Blog: ICO regulatory sandbox. London: Information Commissioner's Office; 2020 (https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/11/sandbox-helps-develop-innovative-tools-to-combat-financial-crime/, accessed 13 February 2021).
243. Fihn SD, Saria S, Mendonça E, Hain S, Matheny M, Shah N et al. Deploying AI in clinical settings. In: Matheny M, Thadaney Israni S, Ahmed M, Whicher D, editors. Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington DC: National Academy of Medicine; 2019.
244. Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ. 2019;5(2):e16048.
245. What is impact assessment. Fact sheet. Fargo (ND): International Association for Impact Assessment; 2009 (https://www.iaia.org/uploads/pdf/What_is_IA_web.pdf, accessed 9 November 2020).
246. Guiding principles on business and human rights: Implementing the United Nations protect, respect, and remedy framework. Geneva: Office of the High Commissioner of Human Rights; 2011 (https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf, accessed 9 November 2020).
247. Human rights impact assessments. National Action Plans on Business and Human Rights. Copenhagen: Danish Institute for Human Rights; 2020 (https://globalnaps.org/issue/human-rights-impact-assessments/, accessed 19 November 2020).
248. French corporate duty of vigilance law. Brussels: European Coalition of Corporate Justice; 2017 (https://corporatejustice.org/documents/publications/french-corporate-duty-of-vigilance-law-faq.pdf, accessed 19 November 2020).
249. Marlow J. New EU law requiring human rights due diligence on the cards for 2021. Blog, 28 July 2020. Paris: Linklaters LLP (https://www.linklaters.com/en/insights/blogs/linkingesg/2020/july/new-eu-law-requiring-human-rights-due-diligence-on-the-cards-for-2021, accessed 19 November 2020).
250. Algorithmic impact assessments: A practical framework for public agency accountability. New York City (NY): AI Now Institute; 2018 (https://www.ftc.gov/system/files/documents/public_comments/2018/08/ftc-2018-0048-d-0044-155168.pdf, accessed 19 November 2020).
251. MacCarthy M. An examination of the Algorithmic Accountability Act of 2019. Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression. Amsterdam: Institute for Information Law; 2019 (https://www.ivir.nl/publicaties/download/Algorithmic_Accountability_Oct_2019.pdf, accessed 19 November 2020).
252. Data protection impact assessment (DPIA). How to conduct a data protection impact assessment (template included). GPPR.EU (https://gdpr.eu/data-protection-impact-assessment-template/, accessed 19 November 2020).
253. Rahwan I, Cebrian M, Obradovich N, Bongard J, Bonnefon JF, Breazeal C et al. Machine behaviour. Nature. 2019;568:477–86 (2019).

254. Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. JAMA. 2019;322(18):1765–6.
255. Price WN. Artificial intelligence in healthcare: Applications and legal implications. The SciTech Lawyer. 2017;14(1). University of Michigan Law School Scholarship Repository (https://repository.law.umich.edu/cgi/viewcontent.cgi?article=2932&context=articles, accessed April 2021).
256. Gerke S, Minssen T, Cohen G. Chapter 12. Ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr A, Memarzadeh K, editors. Artificial intelligence in healthcare. Cambridge (MA): Academic Press;2020:295–336.
257. Ordish J. Briefing. Legal liability for machine learning in healthcare. Cambridge: PHG Foundation; 2018 (https://www.phgfoundation.org/documents/briefing-note-legal-liability-for-machine-learning-in-healthcare.pdf, accessed 22 November 2020).
258. Minssen T, Mimler M, Mak V. When does stand-alone software qualify as a medical device in the European Union? The Cjeu's decision in Snitem and what it implies for the next generation of medical devices. Med Law Rev. 2020;28(3):615–24.
259. Evans BJ, Pasquale FA. Product liability suits for FDA-regulated AI/ML software (Brooklyn Law School, Legal Studies Paper No. 656). In: Cohen IG, Minssen T, Price WN II, Robertson C, Shachar C, editors. The future of medical device regulation: Innovation and protection. Cambridge: Cambridge University Press; 2021 (https://ssrn.com/abstract=3719407, accessed April 2021).
260. Report from the Commission to the European Parliament, the Council, and the European Economic and Social Committee: Report on the safety and liability implications of artificial intelligence, the Internet of Things and robotics. Brussels: European Commission; 2020 (https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1593079180383&uri=CELEX%3A52020DC0064, accessed 15 July 2020).
261. Price WN II. Medical malpractice and black-box medicine (University of Michigan Public Law Research Paper No. 536). In: Cohen IG, Minssen T, Price WN II, Robertson C, Shachar C, editors. The future of medical device regulation: Innovation and protection. Cambridge: Cambridge University Press; 2021 (https://ssrn.com/abstract=2910417, accessed April 2021).
262. Husgen J. Product liability suits involving drug or device manufacturers and physicians: the learned intermediary doctrine and the physician's duty to warn. MO Med. 2014;111(6):478–81.
263. Thomas S. Artificial intelligence and medical liability (Part II). Bill of Health. 10 February 2017. (https://blog.petrieflom.law.harvard.edu/2017/02/10/artificial-intelligence-and-medical-liability-part-ii/, accessed 17 November 2020).
264. No fault compensation in New Zealand: Harmonizing injury compensation, provider accountability, and patient safety. Commonwealth Fund, 24 February 2006 (https://www.commonwealthfund.org/publications/journal-article/2006/feb/no-fault-compensation-new-zealand-harmonizing-injury, accessed 20 March 2021).
265. McNair D, Price WN. Health care AI: Law, regulation, and policy. In: Matheny M, Thadaney Israni S, Ahmed M, Whicher D, editors. Artificial intelligence in health care: The hope, the hype, the promise, the peril. Washington DC: National Academy of Medicine; 2019.
266. Digital health (A/71/A/CONF./1). Seventy-first World Health Assembly. Geneva: World Health Organization; 2018 (https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_ACONF1-en.pdf, accessed 20 November 2020).
267. Elements of informed consent. Seattle (WA): Sage Bionetworks; 2020 (https://sagebionetworks.org/tools_resources/elements-of-informed-consent/, accessed 14 March 2021).
268. Cohen IG. Is there a duty to share healthcare data? In: Cohen IG, Lynch HF, Vayena E, Gasser U, editors. Big data, health law, and bioethics. Cambridge: Cambridge University Press, 2018:209–22.

269. Otake T. Medical big data to be pooled for disease research and drug development in Japan. Japan Times, 15 May 2017 (https://www.japantimes.co.jp/news/2017/05/15/reference/medical-big-data-pooled-disease-research-drug-development-japan/#:~:text=The%20law%2C%20commonly%20called%20Jisedai,the%20development%20of%20new%20drugs, accessed 14 February 2021).
270. Regulations on data governance – questions and answers. Brussels: European Commission; 2020 (https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2103, accessed 13 December 2020).
271. Spector-Bagdady K, Hutchinson R, O'Brien Kaleba E, Kheterpal S. Sharing health data and biospecimens with industry – A principle-driven, practical approach. NEJM. 2020;382(22):2072–5.
272. Greenberg Z. What is the blood of a poor person worth? New York Times, 1 February 2019 (https://www.nytimes.com/2019/02/01/sunday-review/blood-plasma-industry.html, accessed 17 November 2020).
273. Data protection guide. London: Privacy International; 2018 (https://privacyinternational.org/report/2255/data-protection-guide-complete, accessed 14 February 2021).
274. Bowan N. After seven year wait, South Africa's data protection law enters into force. Portsmouth (NH): International Association of Privacy Professionals; 2020 (https://iapp.org/news/a/after-a-7-year-wait-south-africas-data-protection-act-enters-into-force/, accessed 15 February 2021).
275. National Data Guardian: What we do. London: HM Government; 2020 (https://www.gov.uk/government/organisations/national-data-guardian/about, accessed 5 November 2020).
276. Our charter: Tūtohinga. Aukland: Te Mana Raraunga (Maori Data Sovereignty Network); 2020 (https://www.temanararaunga.maori.nz/tutohinga, accessed 5 November 2020).
277. Schnarch B. Ownership, access, control, and possession (OCAP) or self-determination applied to research. Int J Indigenous Health. 2004;1(1) (https://jps.library.utoronto.ca/index.php/ijih/article/view/28934, accessed 19 November 2020).
278. Sharing sensitive health data in a federated data consortium model. An eight-step guide. Insight report. Geneva: World Economic Forum; 2020 (http://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf, accessed 19 November 2020).
279. Creating the right framework to realise the benefits for patients and the NHS where data underpins innovation. London: Department of Health and Social Care; 2019 (https://www.gov.uk/government/publications/creating-the-right-framework-to-realise-the-benefits-of-health-data/creating-the-right-framework-to-realise-the-benefits-for-patients-and-the-nhs-where-data-underpins-innovation, accessed 19 November 2020.
280. Bhunia P. Data futures partnership in New Zealand issues guidelines for organisations to develop social license for data use. Open Government, 27 October 2017 (https://opengovasia.com/data-futures-partnership-in-new-zealand-issues-guidelines-for-organisations-to-develop-social-license-for-data-use/#:~:text=The%20Data%20Futures%20Partnership%20is,control%20data%2Dsharing%20ecosystem.%E2%80%9D, accessed 19 November 2020)
281. WHO data principles. Geneva: World Health Organization; 2021. (https://www.who.int/data/principles, accessed 13 March 2021).
282. WHO data sharing policy: Implementation suggestions. Geneva: World Health Organization; 2020 (https://cdn-auth-cms.who.int/media/docs/default-source/world-health-data-platform/who-data-sharing-policy-implementation-suggestions-10-august-2020.pdf?sfvrsn=fd365554_2, accessed 19 April 2021).
283. Genomic data sharing policy. Bethesda (MD): National Institutes of Health; 2014 (https://www.federalregister.gov/documents/2014/08/28/2014-20385/final-nih-genomic-data-sharing-policy, accessed 12 September 2020)
284. Majumder MA, Guerrini CJ, Bollinger JM, Deegan RC, McGuire AL. Sharing data under the 21st Century Cures Act. Genet Med. 2017;19(12):1289–94.

285. HHS finalizes historic rules to provide patients with more control of their patient data. Washington DC: Department of Health and Human Services; 2020 (https://www.hhs.gov/about/news/2020/03/09/hhs-finalizes-historic-rules-to-provide-patients-more-control-of-their-health-data.html, accessed 12 September 2020)
286. All of Us Research Program. Bethesda (MD): National Institutes of Health; 2020 (https://allofus.nih.gov/, accessed 14 November 2020).
287. European health data space. Brussels; European Commission; 2020 (https://ec.europa.eu/health/ehealth/dataspace_en, accessed 14 November 2020).
288. Digital innovation hub programme prospectus. Appendix: Principles for participation. London: Health Data Research UK; 2019 (https://www.hdruk.ac.uk/wp-content/uploads/2019/07/Digital-Innovation-Hub-Programme-Prospectus-Appendix-Principles-for-Participation.pdf, accessed 15 November 2020).
289. Ornstein C, Thomas K. Sloan Kettering's cozy deal with start-up ignites uproar. New York Times, 20 September 2018 (https://www.nytimes.com/2018/09/20/health/memorial-sloan-kettering-cancer-paige-ai.html, accessed 19 November 2020).
290. The world's most valuable resource is no longer oil, but data. The Economist, 6 May 2017 (https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource, accessed 22 August 2020).
291. Rajan A. Data is not the new oil. BBC News Online, 9 October 2017 (https://www.bbc.com/news/entertainment-arts-41559076, accessed 13 September 2020).
292. Marr B. Here's why data is not the new oil. Forbes Magazine, 5 March 2018 (https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#6a65a1453aa9, accessed 13 September 2020)
293. Hilty R. Big data: Ownership and use in the digital age. In: Seuba X, Geiger C, Penin J, editors. Intellectual property and digital trade in the age of artificial intelligence and big data (Global perspectives and challenges for the intellectual property system. Issue No. 5). Geneva: International Centre for Trade and Sustainable Development; Strasbourg: Center for International Intellectual Property Studies; 2018 (http://www.ceipi.edu/fileadmin/upload/DUN/CEIPI/Documents/Publications_CEIPI___ICTSD/CEIPI-ICTSD_Issue_5_Final.pdf, accessed 24 August 2020).
294. Minssen T, Pierce J. Big data and intellectual property in the health and life sciences. In: Cohen IG, Lynch HF, Vayena E, Gasser U, editors. Big data, health law, and bioethics. Cambridge: Cambridge University Press, 2018.
295. Andanda P. Towards a paradigm shift in governing data access and related intellectual property rights in big data and health-related research. Int Revf Intellectual Property Competition Law. 2019;50:1052–81.
296. Sherkow JS, Minssen T. AIRR data under the EU Trade Secrets Directive – Aligning scientific practices with commercial realities. In: Schovsbo J, Riis T, Minssen T, editors. The harmonization and protection of trade secrets in the EU – An appraisal of the EU Directive. Cheltenham: Edward Elgar Publishing; 2020:239–68.
297. Minssen T, Schovsbo J. Big data in the health and life sciences: What are the challenges for European competition law and where can they be found? In: Seuba X, Geiger C, Penin J, editors. Intellectual property and digital trade in the age of artificial intelligence and big data (Global perspectives and challenges for the intellectual property system. Issue No. 5). Geneva: International Centre for Trade and Sustainable Development; Strasbourg: Center for International Intellectual Property Studies; 2018 (http://www.ceipi.edu/fileadmin/upload/DUN/CEIPI/Documents/Publications_CEIPI___ICTSD/CEIPI-ICTSD_Issue_5_Final.pdf, accessed 22 August 2020).

298. European Open Science Cloud (https://www.eosc-portal.eu/).
299. Corrales Compagnucci, M, Minssen T, Seitz C, Aboy M. Lost on the high seas without a safe harbor or a shield? Navigating cross-border data transfers in the pharmaceutical sector after Schrems II invalidation of the EU-US privacy shield. Eur Pharmaceut Law Rev. 2020;4(3):153–60.
300. Abbott R. Inventive machines: Rethinking invention and patentability. In: Seuba X, Geiger C, Penin J, editors. Intellectual property and digital trade in the age of artificial intelligence and big data (Global perspectives and challenges for the intellectual property system. Issue No. 5). Geneva: International Centre for Trade and Sustainable Development; Strasbourg: Center for International Intellectual Property Studies; 2018 (http://www.ceipi.edu/fileadmin/upload/DUN/CEIPI/Documents/Publications_CEIPI___ICTSD/CEIPI-ICTSD_Issue_5_Final.pdf, accessed 22 August 2020).
301. EPO publishes grounds for its decision to refuse two patent applications naming a machine as an inventor. Munich: European Patent Office, 28 January 2020 (https://www.epo.org/news-events/news/2020/20200128.html, accessed 21 March 2021).
302. Porter J. US Patent Office rules that artificial intelligence cannot be a legal inventor. The Verge, 29 April 2020 (https://www.theverge.com/2020/4/29/21241251/artificial-intelligence-inventor-united-states-patent-trademark-office-intellectual-property, accessed 22 August 2020).
303. Aboy M, Liddell K, Crespo C, Cohen IG, Liddicoat J, Gerke S et al. How does emerging patent case law in the US and Europe affect precision medicine? Nature Biotechnol. 2019;37:1118–26.
304. West DM. The role of corporations in addressing AI's ethical dilemmas. Washington DC: Brookings Institute; 2018 (https://www.brookings.edu/research/how-to-address-ai-ethical-dilemmas/, accessed 24 August 2020).
305. Mittelstadt B. Principles alone cannot guarantee ethical AI. Nature Machine Intelligence. 2019;1:501–7.
306. Metz C, Wakabayashi D. Google researcher said she was fired over paper highlighting bias in AI. The New York Times, 3 December 2020 (https://nyti.ms/2I8oves, accessed 16 December 2020).
307. Cath C. Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Phil Trans R Soc A. 2018;376:20180080.
308. WHO launches a chatbot on Facebook messenger to combat misinformation. Geneva: World Health Organization; 2020 (https://www.who.int/news-room/feature-stories/detail/who-launches-a-chatbot-powered-facebook-messenger-to-combat-covid-19-misinformation, accessed 29 September 2020).
309. Facebook's algorithm: A major threat to public health. Avaaz, 19 August 2020 (https://avaazimages.avaaz.org/facebook_threat_health.pdf, accessed 19 November 2020.
310. Lee D, Murphy H. Facebook accused of failing to tackle medical hoaxes. Financial Times, 20 August 2020 (https://www.ft.com/content/f33f7d61-a8df-40b9-82a8-75f2a41210bc, accessed 24 August 2020).
311. Jin KX. Keeping people safe and informed about the coronavirus. Facebook, 3 December 2020 (https://about.fb.com/news/2020/12/coronavirus/?utm_source=STAT+Newsletters&utm_campaign=01f1d5a35b-health_tech_COPY_01&utm_medium=email&utm_term=0_8cab1d7961-01f1d5a35b-151577169, accessed 16 December 2020).
312. Brodwin E. Facebook's Covid-19 misinformation campaign is based on research. The authors worry Facebook missed the message. STAT News, 1 May 2020 (https://www.statnews.com/2020/05/01/facebooks-covid-19-misinformation-campaign-is-based-on-research-the-authors-worry-facebook-missed-the-message/?utm_source=STAT+Newsletters&utm_campaign=01f1d5a35b-health_tech_COPY_01&utm_medium=email&utm_term=0_8cab1d7961-01f1d5a35b-151577169, accessed 14 December 2020).

313. Isaac M, Wakabayashi D, Cave D, Lee E. Facebook blocks news in Australia, diverging with Google on proposed law. The New York Times, 17 February 2021. (https://www.nytimes.com/2021/02/17/technology/facebook-google-australia-news.html, accessed 24 February 2021).
314. Taylor J. Facebook's botched Australia new ban hits health departments, charities, and its own pages. The Guardian, 18 February 2021 (https://www.theguardian.com/technology/2021/feb/18/facebook-blocks-health-departments-charities-and-its-own-pages-in-botched-australia-news-ban, accessed 24 February 2021).
315. Taylor J, McGowan M, Bland A. Misinformation runs rampant as Facebook says it may take a week before it unblocks some pages. The Guardian, 19 February 2021 (https://www.theguardian.com/technology/2021/feb/19/misinformation-runs-rampant-as-facebook-says-it-may-take-a-week-before-it-unblocks-some-pages, accessed 24 February 2021).
316. International Organization for Standardization. Geneva (https://www.iso.org/home.html).
317. Health level 7 International. Ann Arbor (MI) (http://www.hl7.org/).
318. Brown KV. Alphabet's Verily plans to use big data as health insurance tool. Employee Benefit News, 25 August 2020 (https://www.benefitnews.com/articles/alphabets-verily-plans-to-use-big-data-as-health-insurance-tool, accessed 12 September 2020)
319. Ackroyd AT. Tencent-backed WeDoctor makes IPO appointment in Hong Kong and writes prescription for digital healthcare post-pandemic. South China Morning Post, 4 June 2020 (https://www.scmp.com/business/banking-finance/article/3087385/tencent-backed-wedoctor-makes-ipo-appointment-hong-kong, accessed 28 August 2020).
320. Hello world: Artificial intelligence and its use in the public sector. Paris: Organization for Economic Co-operation and Development; 2019 (https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf, accessed 28 August 2020).
321. The beginning of AI revolution in UAE healthcare. Global Business Outlook, 8 October 2020 (https://www.globalbusinessoutlook.com/the-beginning-of-ai-revolution-in-uae-healthcare/, accessed 5 December 2020).
322. Working document: Enforcement mechanisms for responsible #AIforAll. New Delhi: NITI Aayog; 2020 (https://niti.gov.in/sites/default/files/2020-11/Towards-Responsible-AI-Enforcement-of-Principles.pdf, accessed 12 December 2020).
323. Assessing if artificial intelligence is the right solution. London: HM Government; 2019 (https://www.gov.uk/guidance/assessing-if-artificial-intelligence-is-the-right-solution, accessed 28 August 2020).
324. Committee on Standards in Public Life. Artificial intelligence and public standards: report. London: HM Government; 2020 (https://www.gov.uk/government/publications/artificial-intelligence-and-public-standards-report, accessed 12 February 2020).
325. Martinho-Truswell E. How AI could help the public sector, Harvard Business Review, 29 January 2019 (https://hbr.org/2018/01/how-ai-could-help-the-public-sector, accessed 30 August 2020).
326. Digital technology, social protection and human rights: Report of the United Nations Special Rapporteur for extreme poverty. Geneva: Office of the High Commissioner for Human Rights; 2019 (https://www.ohchr.org/EN/Issues/Poverty/Pages/DigitalTechnology.aspx, accessed 21 March 2021).
327. Derrington D. Artificial intelligence for health and healthcare. McLean (VA): The MITRE Corporation; 2017 (https://www.healthit.gov/sites/default/files/jsr-17-task-002_aiforhealthandhealthcare12122017.pdf, accessed 17 August 2020)
328. Federal Trade Commission. Twitter; 2020 (https://twitter.com/FTC/status/1285578871803437057, accessed 15 March 2021).

329. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Guardian, 17 March 2018 (https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election, accessed 23 November 2020).
330. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med. 2018;15(11):e1002689.
331. AI Policy Observatory fact sheet. Paris: Organization for Economic Co-operation and Development; 2020 (https://www.oecd.org/going-digital/ai/about-the-oecd-ai-policy-observatory.pdf, accessed 17 November 2020).
332. Rao AS, Verweij G. Sizing the prize: What's the real value of AI for your business and can you capitalise? London: Pricewaterhouse Coopers; 2019 (https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf, accessed 24 August 2020)
333. Davis SLM. Perspective. The Trojan horse: Digital health, human rights, and global health governance. Health Human Rights J. 2020;22:41–8.
334. Artificial intelligence: Canada and France work with international community to support the responsible use of AI. Paris: Government of France; 2019 (https://www.gouvernement.fr/en/artificial-intelligence-canada-and-france-work-with-international-community-to-support-the, accessed 24 August 2020.)
335. The Global Partnership on Artificial Intelligence (https://gpai.ai/).
336. Report of the Secretary-General: Roadmap for digital cooperation. New York City (NY): United Nations; 2020 (https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf, accessed 17 September 2020).

ANNEX. CONSIDERATIONS FOR THE ETHICAL DESIGN, DEPLOYMENT AND USE OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR HEALTH

The following provides practical guidance for several key groups that use AI in the health field: AI designers and developers, ministries of health and health care institutions and providers. It reflects the main principles, ideas and recommendations in this report.

A1. Considerations for AI developers

The following considerations are for individuals, research organizations and companies involved in the design, deployment and updating of AI technologies used in health. AI developers include professionals with expertise in computer science or AI, who often also have a background in clinical or health care. Some AI developers are not sited in health systems, even though the products they design will play an increasingly important role in health. Some providers and hospitals are investing in and designing AI technologies and should consider the issues listed below with their existing ethical obligations as medical providers.

Developers, research organizations and companies should consider systems to ensure that the values, principles and processes that guide their operations are aligned with the expectations of health systems.

The considerations listed below are not comprehensive but are steps that developers and companies should take to ensure that the technologies they design and deploy are used for the benefit of patients and providers. Three areas should be considered: the design, development and deployment of an AI technology, with further consideration of improving it after deployment.

Designing an AI technology

1. Clarify the objectives

An AI technology or tool can be used alone or as an integral part of a system. The intended uses, the values and the indirect outcomes for users should be clearly defined.

Specific considerations:

- Define the intended uses and the expected outcomes.
- What are the main functions of the tool?
- Who will use the tool?
- How will it be used
- When and where will it be used or not used?
- Will there be secondary (indirect) users?
- How should the objectives and functions be prioritized according to the available resources?

- Will use of the tool have indirect outcomes?
- Are the validity and efficiency of the tool limited over time?

2. Engage multiple stakeholders and understand contexts.
AI technologies used in health care depend on the context and must be designed to work appropriately for different types of health-care providers and different uses by patients or practitioners before, during or after clinical care.

Specific considerations:

- Define all possible contexts in which the AI technology will be used, including geographical scope, users' background and main languages, digital skills and regulatory frameworks.
- Involve individuals who understand various contexts in design to align the objectives and expected outcomes and avoid transferring bias from the data and amplifying it.
- Design, discuss and validate the formulation, conceptualization, proposed approach and solution with stakeholders in the targeted settings, including policy- and decision-makers, project owners and leaders, project managers, solution engineers and developers, potential users, domain experts and experts in ethics and information privacy.
- Clearly delineate responsibilities during design, development and deployment and the conditions to be fulfilled for attribution of responsibility.
- Determine the operational and technical limitations to designing, developing, testing, using and maintaining the tool, including human resources, expertise and software and hardware requirements.

3. Define relevant ethical issues through consultation.
Each AI technology will require consideration of ethical issues, such as bias, privacy, data collection and use and human autonomy (among the principles listed in section 5 of this report). Ethical concerns that often emerge during consultation should be identified and integrated into the design and development. (Recommendations for addressing bias and privacy, two ethical issues that are often relevant for the design of AI technologies for health, are discussed below.)

4. Assess risks.
Risk assessment and mitigation are necessary in the design and development of technologies for use in human health. Risk should be assessed at each stage of development and reassessed regularly with stakeholders. The aim of developers should be for the AI technology to achieve the intended outcomes with a reduced level of risk. All major trade-offs should be clearly identified and considered.

Specific considerations:

- What are the expected outcomes?
- What are the potential secondary and unexpected outcomes?

- What would be the impact and consequences of the unexpected outcomes?
- What are the available resources and potential trade-offs?
- What approaches would mitigate risk?

5. Address biases.

Biases in data due to past or continuing discrimination could be replicated. An AI technology should be used only if such bias can be mitigated. AI should be designed to reduce inequities and bias.

Specific considerations:

- Determine how the study data were collected and how new study data will be collected, and look for any bias in the data according to the context.
- Consider the majority and minority groups included in the data and whether any under-representation that results in bias can be mitigated.
- Examine the effects of ethnicity, age, race, gender and other traits, and ensure that AI technologies with biases do not have negative impacts on individuals and groups according to these different characteristics.
- Prepare effectively and demonstrably for post-implementation surveillance of the application.

6. Privacy by design and privacy by default

All possible steps should be taken to safeguard the security, privacy and confidentiality of the information used to develop and validate an AI technology in relevant contexts and of the information and data collected and produced by the AI technology.

Specific considerations:

- Map the possible vulnerability of an AI technology with respect to privacy and reverse engineering in context.
- Identify data protection vulnerabilities in contracts and collaborations with (other) commercial parties and data-sharing systems and networks.
- Select design options that favour privacy and ensure that any reduction in privacy is consciously agreed to.
- Safeguard data protection and privacy preservation over time and with technology updates.

Developing an AI technology

1. Identify regulatory requirements.

Regulatory frameworks for AI are evolving. While most regulatory frameworks address data protection, data security and privacy, emerging governance guidelines include equal access and human autonomy. Compliance measures should be included in development and updates of a technology.

Specific considerations:

- Adhere to country-specific or regional export rules and guidelines, such as the EU GDPR, Singapore's Personal Data Protection Act or the US Health Insurance Portability and Accountability Act.
- Identify open concepts and open norms that should be specified for compliance, e.g. in GDPR Article 22, the "far reaching effects" in "Person may not be subjected solely to automated decision procedure with far reaching effects".
- Define relevant open norms and concepts that can be justified to affected parties and experts with relevant knowledge of the application.

2. Establish data management plans.
Clear management plans and protection guidelines should be established for data collection, storage, organization and access to ensure data security and safeguard privacy and confidentiality.

Specific considerations:

- Understand the data collection and sharing requirements and regulations in the countries, sectors and institutions of potential users, including legal requirements for managing consent for the use of training data.
- Determine the type of data that are being collected and where and how the data will be stored.
- Assess the physical infrastructure and operational processes that can be used to ensure data security and integrity.
- Understand and determine how confidentiality and privacy will be protected in different contexts.
- Establish guidelines and protocols for proper collection, storage, organization, access and use of personal, proprietary and public data in different contexts.
- Determine how long the data will be stored, when the data could be shared and other temporal considerations.
- Give preference to the use of anonymized data whenever possible.
- Determine who is responsible for data governance and ensure appropriate follow-up.
- Clearly identify all groups who will have access to the data throughout the product's life cycle.
- Determine any type of secondary use of data that could be allowed.

3. Adopt standards and best practices.
Ensure the compliance and/or interoperability of the AI technology with other technologies that will be introduced into health systems. One or more established international, regional or national standards and/or performance benchmarks for an AI technology should be adopted according to regulations, guidance and application requirements, design and development plans.

Specific considerations (examples of standards):

- ISO standards (security and privacy)
- US National Institute of Standards and Technology (security and privacy)
- IEEE 7000 series (privacy and fairness)
- Health Level 7 (transfer of administrative and clinical health data)

Deploying an AI technology and improving it after deployment

1. Engage and educate multiple stakeholders for deployment and maintenance.
Prioritize inclusivity throughout to ensure better understanding of needs and to build adapted solutions for multiple stakeholders.

Specific considerations

- Clearly delineate responsibility for what to do, when and how.
- Design, discuss and validate the proposed approach with various stakeholders in all targeted regions, including policy- and decision-makers, project owners and leaders, project managers, solution engineers and developers, potential users, domain experts and experts in ethics and information privacy.
- Train stakeholders in why, how and when to use the tool, including the main objectives, functions and features and differences among usage scenarios, when applicable.
- Engage continuously with stakeholders, and support users.

2. Evaluate and improve performance.
The outcomes and impact on health care of the AI technology should be assessed formally, and the design and development of the technology continuously improved according to the ethical principles that initially guided its development and to new governance guidelines and all applicable legal obligations and regulations. The risks of the technology and of its intended usage in different health care settings should be assessed regularly to manage its deployment, continuous development and maintenance.

Specific considerations

The accuracy and risks of error of the AI technology should be evaluated to assess implications for:

- Incorporating, verifying and validating changes to the tool or system;
- monitoring and ensuring the effectiveness and usefulness of the tool or system over time;
- how long the results or the technology can be used;
- how often the tool or system should be updated; and
- who is responsible for updating.

A2. Considerations for ministries of health

The following considerations are intended for ministries of health, which will have the primary responsibility for determining whether and how AI technologies should be integrated into health systems, the conditions under which they should be used, the protection of individuals that must accompany use of such technologies and policies that can address both expected and unexpected ethical challenges. Evaluation, regulation, deployment and oversight of AI technologies will require inter-ministerial coordination. Thus, while these considerations are directed to ministries of health, implementation will require collaboration with other relevant ministries, such as of information technology and education.

These considerations are not comprehensive but may be a starting-point for ministries of health to ensure that the use of AI technologies is consonant with the wider objective of the government to provide affordable, equitable, appropriate, effective health care, with the goal of attaining universal health coverage. Three areas should be considered: how ministries should protect the health and safety of patients, how they should prepare for the introduction and use of AI technologies and how they should address ethical and legal challenges and protect human rights.

How to protect the health and safety of patients

1. Assess whether AI technologies are appropriate and necessary.

AI technologies should be used only if they are necessary and appropriate and contribute to achieving universal health coverage. They should not divert attention and resources from proven but underfunded interventions that would reduce morbidity and mortality.

Specific considerations

- Evaluate the institutional and regulatory context and infrastructure to determine whether the technology would be as cost–effective as "traditional" technologies and whether its introduction and use are in accordance with human rights.
- Conduct an impact assessment before deciding whether to implement or continue use of AI in the health system.
- Calculate the risk–benefit ratio of adoption, investment and uptake of an AI technology, and make the information available to stakeholders so that they can provide input to any evaluation or decision.
- Manage the ethical challenges of the AI technology (e.g. equitable access, privacy) appropriately.

2. Testing, monitoring and evaluation

AI must be rigorously tested, monitored and evaluated. Clinical trials can provide assurance that any unanticipated hazards or consequences of AI-based applications

are identified and addressed (or avoided entirely) and an approved AI device can be re-tested and monitored to measure its performance and any changes that may occur once it has been approved.

Regulatory agencies can support testing, transparent communication of outcomes and monitoring of' the performance and efficacy of a technology. Many LMIC still lack sufficient regulatory capacity to assess drugs, vaccines and devices, and the rapid arrival of AI technologies could mean that their regulatory agencies cannot accurately assess or regulate such technologies for the public good.

Specific considerations

- Countries should have sufficient regulatory capacity to ensure rigorous scrutiny of AI technologies on which countries rely in health care.
- For certain low-risk AI technologies, regulators may consider "lighter" premarket scrutiny.
- AI technologies should be tested prospectively in randomized trials and not against existing laboratory datasets.
- Regulatory scrutiny should be applied when data from non-health devices are imputed and used to train AI health technologies.

3. Assign liability.
Reliance on AI technologies entails responsibility, accountability and liability and also compensation for any undue damage.

Specific considerations

- Ministry of health experts should evaluate AI tools to ensure accountability for any negative consequences that arise from their use.
- Liability rules used in clinical care and medicine should be modified to assess and assign liability, including product liability, the personal liability of decision-makers, input liability and liability to data donors. The rules should include causal responsibility, objective liability regimes and liability for retrospective harm as well as mechanisms for assigning vicarious liability when appropriate.

4. Ensure that all people are guaranteed redress in the legal system.
Processes should be available for compensation of undue damage caused by use of AI technologies.

Specific considerations

- Independent oversight should be available to ensure equitable access to health care of appropriate quality.
- Swift, accessible mechanisms should be available for complaint, including for patients and health staff to demand protection of personal data and particularly of sensitive health data.

Prepare for the introduction and use of AI technologies.

1. Institutional preparedness and technical capacity

Ministries of health should have the necessary human and technical resources to realize the full benefits of AI technologies for health while mitigating any negative impacts.

Specific considerations

- Training and capacity-building based on established criteria should be organized for government officials to evaluate whether an AI technology is based on ethical principles.
- Health-care authorities and medical professionals should be involved and engaged in AI design and, when possible, software engineering.
- Civil society, medical staff and patient groups should be consulted about the introduction of AI technology and included in both external audit and monitoring of its functioning.
- The introduction of an AI technology should be accompanied by appropriate investments by the health system to capture its benefits. For example, tools to predict a disease outbreak should be complemented by robust surveillance systems and other measures to respond effectively to an outbreak.

2. Infrastructure for AI technologies

The right infrastructure is a prerequisite for proper deployment of AI in a health-care system.

Specific considerations

- Criteria should be established to identify and measure the infrastructure requirements, including for operation, maintenance and oversight.
- When necessary, infrastructure should be provided or strengthened with civil society support and international cooperation.
- Ministries of health should identify effective alternatives if any infrastructure is lacking, if the AI technology is too expensive or if it poses a high risk to patients.

3. Management of data

Data must be of high quality to prevent unintended harm from use of AI systems, as limited, low-quality or inaccurate data could result in biased inferences, misleading data analyses and poorly designed applications for health. Other critical elements of health data management include protecting the privacy and confidentiality of patient data and the rules for sharing such data.

Specific considerations

- Data processing (including from non-medical devices) and its representativeness, accuracy, harmonization, accessibility, interoperability and reusability should be

regulated, with the informed consent of data providers (patients).
- Access to and use of data from digital self-care applications and/or wearable technologies should also be regulated. Data from these applications and technologies should be collected, stored and used in accordance with principles for data minimization.
- Patients and consumers who provide data should have access to and be allowed to reuse and thereby benefit from their data. Their data should not be monopolized by an AI technology provider.
- Quality control measures should be implemented to ensure the representativeness of data from different population groups.
- Mechanisms and procedures should be in place to collect relevant patient data to train AI technology according to the environment, culture and specifics of the community in which the technology is intended to be used.
- Patients and consumers should know what data are used in training AI systems.

Address ethical and legal challenges and protect human rights.

1. Preserve and enhance human autonomy.

AI technologies for health should enhance human decision-making and empower medical professionals (clinicians and providers) rather than replace them.

Specific considerations

- Human judgement should be used with regard to prediction of disease and/or recommended treatment by an AI technology.
- Ministries of health should designate the types of information with which a clinician should be provided to make an independent judgement about an AI result or outcome.
- Meaningful, clear information should be provided to patients to allow them to make informed decisions about health recommendations based on AI technology.

2. Patient agency with regard to predictive algorithms

Use of AI predictive analytics in health care raises ethical concern with respect to informed consent and individual autonomy in decisions about patient and consumer health.

Specific considerations

- The need for an AI technology should be assessed, with the risk of the technology to patient autonomy and well-being.
- Patients should be allowed to refuse AI technologies for health.
- A mechanism should be available to inform patients of the benefits, risks, value, constraints, novelty and scope of an AI tool.

3. Privacy, confidentiality and informed consent in the collection and use of patient data
The autonomy and trust of patients who provide data are paramount, especially meaningful individual control over data. Health-data processing should include respect for the right to privacy and should ensure that patients maintain control over decisions, including their informed consent.

Specific considerations:

- Up-to-date data protection and confidentiality laws should be a prerequisite for use of AI.
- Independent oversight and other forms of redress should be available to protect patient privacy and data confidentiality.
- Data protection supervisory agencies should have sufficient resources for effective privacy protection.
- Ministries of health should employ experts to determine whether AI tools meet standards of privacy to foster the general trust of patients who provide data.
- Ministries of health should have a protocol for collecting, storing and sharing personal data or data that could be identified and ensure that the data are managed in such a way as to protect privacy, including confidentiality and informed consent.
- Ministries of health should ensure that patients have the right to refuse data collection by and the data-sharing requirements of an AI technology. Explicit consent should be given for secondary uses of health data.
- Ministries of health should limit the collection of data to those required and not collect additional data.
- Ministries of health should provide training for health staff in the implications for the human rights of patients as part of capacity-building for use of AI technology.

4. Transparency of AI technologies for health
AI technologies must be provided and relied on transparently in order to assign responsibility and ensure trust and protection of patient rights.

Specific considerations:

- Ministry of health experts should transparently evaluate an AI technology developed by others and make the results of such assessments publicly available throughout the life-cycle of the AI system.
- Ministries of health should ensure that clinicians can explain how an AI system has been validated to patients and their families.
- External experts should have enough information about the AI system and its training data to make independent assessments.

5. Ensure equitable access to AI technologies and related health care.
When an AI technology is considered necessary (see above), ministries of health have an ethical obligation to ensure equitable access to that technology. Diagnostic use of

AI should be extended carefully to avoid situations in which large numbers of people receive an accurate diagnosis of a health condition in the absence of appropriate treatment options.

Specific considerations

- Ministries of health have a duty to ensure equitable access to all to AI-based health care, regardless of gender, geography, ethnicity and other conditions.
- Ministries of health have a duty to provide treatment after AI-based testing and confirmation of disease.
- Ministries of health should ensure that the benefits of data from AI are fairly shared with the patients who provided the data for AI training and not monopolized by technology service providers.

A3. Considerations for health-care institutions and providers

The following considerations are intended for health-care institutions and providers, such as hospitals, doctors and nurses. While programmers may be those primarily responsible for the design of AI technologies and ministries of health and regulatory agencies for approval and selection of such technologies for use, health-care providers determine which technologies to use and how and may also provide direct feedback to the health-care system, the medical community and the designers of the technologies about whether they meet the needs of patients.

The following is not comprehensive but may be used as a starting point as health-care providers increase use of AI for health care. Use of AI technologies for health outside regular health-care settings is discussed in section 3.1 of the report. Three areas are considered: whether the AI technology is necessary and appropriate; whether the context in which the AI technology will be used is appropriate; and whether a health-care provider should use a particular AI technology.

Is the AI technology necessary and appropriate?

1. Prioritize safety.

Use of AI technology in health care will inadvertently address and could amplify risk-prone decisions, procedures or both. Technology-related risks must be counteracted by risk mitigation strategies, which should be integrated into AI decision-making or be applicable to AI decisions.

2. Promote transparency.

Introduction of any AI technology must be sufficiently transparent that it can be criticized, by the public or by internal review mechanisms.

Specific considerations:

- The source code should be fully disclosed.
- Algorithms must be open to criticism by an in-house or other appropriate expert.
- The data used to train the algorithm, whether certain groups were systematically excluded from such data, how the training data were labelled and by whom (including expertise and appropriateness of labelling) should be known.
- The underlying principles and value sets used for decision trees should be transparent.
- The learned code should be available for independent audit and review by appropriate third parties.

3. Address bias.

Bias due to past or continuing discrimination could be replicated. An AI technology should be used only if such bias can be mitigated, and AI should be designed to reduce inequity and bias.

Specific considerations:

- Ensure that AI with certain biases does not have negative impacts according to race or ethnicity or that the bias can be mitigated.
- If bias cannot be removed, ensure that this is stated transparently and reflected in decisions, e.g. to be taken into consideration by a provider or patient.

4. Safeguard privacy

Health-care providers must prevent re-identification, especially for datasets that can be linked by third parties to re-identify individuals.

Specific considerations

- Understand issues related to privacy and reverse engineering.
- Ensure that any option for use of an AI technology in a clinical setting favours privacy and that any reduction in privacy is actively agreed.
- Take the necessary measures to prevent leakage of identifiable information.

5. Institute regular challenge and review.

Even if an AI technology is deemed appropriate up front, it must be subject to regular challenge and review. This may be necessary due to software erosion, changes in context over time and changes in the AI technology itself as it continues to learn from new data and evolves.

Specific considerations:

- Establish regular technical review, including external review.
- Review whether the AI is having the intended impact, is filling a gap in need and is improving health care.

Is the context in which the AI technology will be used appropriate?

1. Assess whether the AI technology is necessary and appropriate in each clinical setting.

Specific considerations:

- Determine whether the AI technology offers advantages over what is currently offered and fills a gap.
- Compare the risks and benefits of the AI technology with those of current technology.
- Ensure that the AI technology is necessary and the problem is clearly stated to ensure effective delivery of care that justifies use of the technology.
- Ensure that the AI technology is based on sufficient electronic health data.
- Ensure that the health data used were acquired in an ethical manner.
- Ensure the necessary infrastructure for use of the AI technology.

- Confirm the support of experts, including partnerships with academic institutions and commercial entities, and appropriate agreements with respect to IP, accountability, confidentiality, ethics, access and commercialization.
- Establish commonly agreed ethical principles for the collection, sharing and use of the data and its governance.

2. Understand local perspectives.
The perspectives of local consumers should be recognized, particularly the sovereignty of indigenous peoples over their data for the collective benefit of people. This includes determining whether the health service has a "social license" to use AI, i.e. the consent of communities and/or individuals.

Specific considerations:

- Public and consumer communication and education about AI should be adequate.
- Providers should secure a "social license" from the communities involved.
- Providers should ensure sovereignty and governance of indigenous populations over their data.

Should a health-care provider use the AI technology?

1. Ensure that the information provided by an AI technology can be interpreted.
The information derived by an AI technology must be interpreted by a clinician. Human judgement is critical, and the context is important. Clinicians should be able understand the data and variables so that they can explain the principles of the AI application to themselves, colleagues, patients and families.

2. Understand the level of risk.
Decisions made by clinicians on the basis of an AI technology must be transparent and based on understanding that they are appropriate or commensurate with any risk. AI should be used in prevention, treatment, rehabilitation and/or palliative care only if it the risk–benefit ratio is positive. It should not be used if the influence of the technology on risk is unclear or if it could increase or exacerbate risk. Specific guidelines for medical research involving human beings must be followed if AI technology is used experimentally.

3. Ensure responsible use of AI.
Health-care providers must not only ensure that an AI technology is technically accurate but also consider whether it can be used responsibly. Health-care providers should state specifically why AI is appropriate in a particular situation.

Health Ethics and Governance Unit
Research for Health Department

Digital Health and Innovation Department
Division of the Chief Scientist

World Health Organization
Avenue Appia 20
1121 Geneva 27
Switzerland